ISSN: 0128-7680

Pertanika Journal of

SCIENCE & TECHNOLOGY

VOLUME 10 NO.1
JANUARY 2002

A scientific journal published by Universiti Putra Malaysia Press

Pertanika Journal of Science & Technology

About the Journal

Pertanika, the poineer journal of UPM, began publication in 1978. Since then, it has established itself as one of the leading multidisciplinary journals in the tropics. In 1992, a decision was made to streamline Pertanika into three journals to meet the need for specialised journals in areas of study aligned with the strengths of the univeristy. These ara (i) **Pertanika Journal of Tropical Agricultural Science, (ii) Pertanika Journal of Science and Technology (iii) Pertanika Journal of Social Science and Humanities.**

Aims and Scope

Pertanika Journal of Science and Technology welcomes full papers and short communications in English or Bahasa Melayu in the fields of chemisty, physics, mathematics and statistics, engineering, environmental control and management, ecology and computer science. It is published twice a year in January and July.

Articles must be reports of research not previously or simultaneously published in other scientific or technical journals.

Communications are notes of a significant finding intended for rapid publication. It should not exceed five double-spaced typewritten pages and must be accompanied by a letter from the author justifying its publication as a communication.

Reviews are critical appraisals of literature in areas that are of interest to a broad spectrum of scientists and researchers. Review papers will be published upon invitation.

Submission of Manuscript

Three complete clear copies of the manuscript are to be submitted to

The Chief Editor
Pertanika Journal of Science and Technology
Universiti Putra Malaysia
43400 UPM Serdang, Selangor Darul Ehsan
MALAYSIA
Tel: (603) 89468855 Fax (603) 89416172

Proofs and Offprints

Page proofs, illustration proofs, the copy-edited manuscript and an offprint order form will be sent to the author. Proofs must be checked very carefully within the specified time as they will not be proofread by the Press editors.

Authors will receive 20 offprints of each article. Additional copies can be ordered from the Secretary of the Editorial Board by filling out the offprint order form.

Published by Universiti Putra Malaysia Press
ISSN No.: 0128-7680

Pertanika Journal of Science & Technology

Volume 10 No. 1, 2002

Contents

Pertanika J. Sci. & Technol. 10(1): 1-12 (2002)
ISSN: 0128-7680

Hydraulic Simulation of Flood Occurrences in a Tropical River System: the Case of Linggi River System

Salim Said, Thamer A. Mohammed, Mohd. Zohadie Bardaie, & ShahNor Basri
Department of Biological & Agricultural Engineering
Faculty of Engineering
Universiti Putra Malaysia
43400 UPM, Serdang, Selangor, Malaysia

Received: 4 April 2000

ABSTRAK

Ramalan banjir adalah penting kerana ianya dapat menolong mengurangkan kerosakan akibat banjir terutamanya di kawasan hilir sungai. Peningkatan dalam kaedah-kaedah numerikal dan teknologi komputer telah menghasilkan banyak model matematik yang boleh digunakan untuk simulasi hidraulik banjir. Simulasi hidraulik banjir dalam satu sistem sungai kebiasaannya termasuk ramalan luas dan kedalaman banjir di sepanjang sungai tersebut. Maklumat sedemikian adalah penting kerana ianya dapat membantu jurutera untuk memberi perhatian yang lebih ke atas reka bentuk bagi mengurangkan kerosakan oleh banjir. Model hidraulik yang digunakan bagi mensimulasikan aliran sungai boleh diklasifikasikan kepada model hidraulik dinamik dan model hidraulik statik. Model HEC-2 statik telah digunakan untuk membuat ramalan hidraulik aliran Sungai Linggi di kawasan Seremban, Malaysia. Model HEC-2 adalah berdasarkan kepada penyelesaian numerikal bagi persamaan tenaga satu dimensi. Kalibrasi dan verifikasi tentang profil permukaan air telah diperolehi dengan menggunakan data rekod bagi Sungai Linggi. Selepas kalibrasi, model HEC-2 telah digunakan untuk meramalkan profil permukaan air untuk Q_2, Q_{10} dan Q_{100} di sepanjang sistem Sungai Linggi. Hasil perbandingan tersebut menunjukkan tiada perbezaan ketara, dengan ralat absolut maksimum adalah 100 mm dan ralat absolut minimum adalah 20 mm sahaja.Bacaan-bacaan ini menunjukkan perbezaan yang kurang daripada 5%.

ABSTRACT

Flood forecasting is important because it can help in reducing the consequences of flood damage especially at the downstream end. Advances in numerical methods and computer technologies, have resulted in the development of many mathematical models which can be used for hydraulic simulation of flood. These simulations usually include the prediction of the extent of flood and its depth along a river system. Information obtained from the simulated models are essential because it can help engineers to take precautionary measures in designing their hydraulic structures. Hydraulic models that are used in the simulation can be classified into dynamic hydraulic models and static hydraulic models. The HEC-2 static hydraulic model was used to predict the flow of Linggi river in the vicinity of Seremban town. HEC-2 model is based on the numerical solution of a one-dimensional energy equation of the steady gradually varied flow using an iteration technique. Calibration and verification of the HEC-2 model were conducted using the recorded data for the Linggi river. After calibration, the model was used to predict the water surface profiles for Q_2, Q_{10}, and Q_{100} along the watercourse of the Linggi river. The predicted water surface profiles were found to be in agreement with the recorded water surface profiles, whereby the maximum computed value of the absolute error in the predicted water surface profile was found to be 100 mm while the minimum absolute error was found to be 20 mm only. In term of percentage, these errors represent a difference of less than 5% between the readings of the computed simulation to the actual field records. Testing process showed that HEC-2 model is sensitive to value of Manning coefficient of roughness and the intervals of cross sections long studied river system.

Keywords: Simulation, modelling, flood, tropical river system, water surface profile

INTRODUCTION

A flood is any abnormally high water-stage or flow over land, in a stream, floodway, lake, or coastal areas that can cause detrimental effects to life and property. Flooding in the river system of a tropical region is mainly due to excessive rainfall in the upper catchment of the river system. This is made worse if the catchment area is infringed with development. The worst flood in Malaysia was recorded in 1926 which had been described as having caused the most extensive damage to the natural environment. Subsequent major floods were recorded in 1931, 1947, 1954, 1957, 1967, and 1971. Floods of lesser magnitude also occurred in 1973, 1979 and 1983 (Ann 1994). As a result of advances in numerical methods and computer technologies, many mathematical models have been developed and used for hydraulic simulation of the flood. The hydraulic simulation of the flood in a river system usually includes the prediction of the extent of flood and its depth along a river course. This type of information is essential because it will help engineers to take precautionary measures in their design so as to minimize the potential flood damage especially at the downstream end. Hydraulic models that are used in the simulation can be classified into dynamic hydraulic models and static hydraulic models. This classification was based on the concept and the approach used in the formulation of these models. An example of a recent work on the static hydraulic model for computing water surface profile in prismatic and non-prismatic channels was developed by Ishikawa (1984). Examples of dynamic hydraulic models were those developed by Lyness and Myers (1994), Molls and Chaudhary (1995), and Sturm and Sadiq (1996). Nik (1996) applied both HEC-2 static hydraulic model and MIKE 11 dynamic hydraulic model to predict the water surface elevation of the Klang River, and found that there was a difference of about 5% in the results obtained. In the present study, the HEC-2 static hydraulic model was calibrated, verified and then applied to predict the water surface profiles along the watercourse of the Linggi river and its tributaries.

MODEL FORMULATION

The hydraulic simulation of the flow in a river, stream or a drain is useful for many water resources projects. Knowledge about the water surface profile in nonprismatic channels is important specifically for flood plain management, flood mitigation and for analysis and design of the river crossing. In this study, the flow along nonprismatic channel was hydraulically simulated and a final form of the mathematical model was used to predict the water surface profile of the Linggi river watercourse. A hydraulic model known as HEC-2, developed by the Hydrologic Engineering Center, USA was used to simulate the flow. The model is based on the numerical solution of a one-dimensional energy equation that can be applied for the flow between two sections of a river reach. In the HEC-2 model, both major and minor losses in energy occurred in a river reach, were considered since these two types of energy losses are effective. Thus, the total energy loss for the flow in a river is due to the friction loss, eddy loss and any other possible minor losses. To explain the mathematical algorithm, it is convenient to refer to the water surface for a natural channel above a datum at the two sections as shown in *Fig. 1*. When the energy principles are applied for the two sections, the following equations were obtained:

$$h_f = \left(\frac{L_l Q_l + L_c Q_c + L_r Q_r}{Q + Q + Q} \right) \left(\frac{Q_{T1} + Q_{T2}}{K_{T1} + K_{T2}} \right) \tag{1}$$

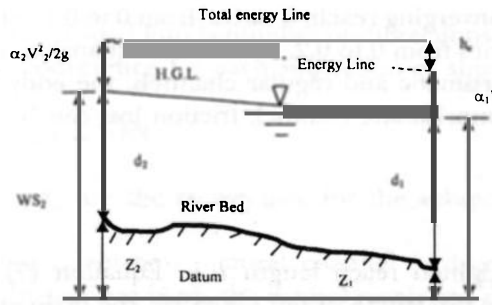


Fig. 1: Natural river reach used in the derivation of HEC-2 model

and $WS_1 = d_1 + z_1$ (2)

$$WS_2 = d_2 + z_1 \tag{3}$$

Substituting Equation (2) and Equation (3) into Equation (1) results in the following equation:

$$WS_2 + \frac{\alpha_2 V_2^2}{2g} = WS_1 + \frac{\alpha_2 V_1^2}{2g} \tag{4}$$

Where WS_1, WS_2 are water surface elevations from a datum for section 1 and section 2 respectively, d_1, d_2 are water depths at section 1 and section 2 respectively, z_1, z_2 are the channel bed elevations above a datum at section 1 and section 2 respectively, V_1, V_2 are average velocities (total discharge per total area of the flow) at section 1 and section 2 respectively, α_1, α_2 are velocity weighting coefficients at section 1 and section 2 respectively, g is the acceleration due to gravity, and he is the energy loss in the reach.

Chow (1959) defined the energy loss in the reach of a river as a combination of the friction loss and eddy loss:

$$h_e = h_f + h_l \tag{5}$$

where h_f is the friction loss and h_l is the eddy loss.

The eddy loss h_l is appreciable in nonprismatic channels and there is no available rational method of evaluating this loss. The eddy loss depends mainly on the velocity head change and may be expressed as shown below:

$$h_l = \theta\left(\alpha_2 \frac{V_2^2}{2g} - \alpha_1 \frac{V_1^2}{2g}\right) \tag{6}$$

where θ is the eddy loss coefficient.

For a gradually converging reach, θ varies from 0 to 0.1, and for a gradually diverging reach θ varies from 0 to 0.2. For abrupt expansions and contractions, θ is about 0.5. For prismatic and regular channels, the eddy loss is practically zero, ($\theta = 0$). For nonprismatic channel, friction loss can be described by the following formula:

$$h_f = L\bar{S}f \tag{7}$$

The discharge-weighted reach length L in Equation (7) is computed by proper proportioning the length of left overbank, the main channel, and right overbank with their respective flows at the end of the reach as given by the following equation :

$$L = \frac{L_l\overline{Q}_l + L_c\overline{Q}_c + L_r\overline{Q}_r}{\overline{Q}_l + \overline{Q}_c + \overline{Q}_r} \tag{8}$$

A representative friction slope S is expressed as follows:

$$\bar{S}f = \left(\frac{Q_{T1} + Q_{T2}}{K_{T1} + K_{T2}}\right)^2 \tag{9}$$

where L_l, L_c, L_r are reach lengths specified for flow in left over bank, main channel and right overbank, respectively, Q_l, Q_c, Q_r are arithmetic average of the flows at the ends of the reach for left overbank, main channel, and right overbank respectively, Q_{T1}, Q_{T2} are the values of the total discharge at section 1 and 2 respectively, K_{T1}, K_{T2} are the composite or total conveyance for section 1 and 2 respectively.

By substituting Equation (8) and Equation (9) into Equation (7), the following equation is obtained:

$$h_f = \left(\frac{L_l Q_l + L_c Q_c + L_r Q_r}{Q_l + Q_c + Q_r}\right)\left(\frac{Q_{T1} + Q_{T2}}{K_{T1} + K_{T2}}\right) \tag{10}$$

Total energy loss in a river reach h_e can be obtained by substituting Equation (10) and Equation (6) into Equation (5):

$$h_f = \left(\frac{L_l Q_l + L_c Q_c + L_r Q_r}{Q_l + Q_c + Q_r}\right)\left(\frac{Q_{T1} + Q_{T2}}{K_{T1} + K_{T2}}\right)^2 + \theta\left(\frac{\alpha_2 V_2^2}{2g} - \frac{\alpha_1 V_1^2}{2g}\right) \tag{11}$$

The total conveyance of a river section can be described as below:

$$K_{Ti} = \frac{Q_{Ti}}{\sqrt{S_i}} \qquad (i = 1, 2, 3, ..., N) \tag{12}$$

If the river section is divided into N number of subsections, the total conveyance is the sum of the conveyance for each subsection as shown below:

$$K_{Ti} = k_1 + k_2 + k_3 + ... + kN \tag{13}$$

where k_1, k_2, k_3, ..., k_N are the conveyance for the subsection number 1, 2, 3, ..., N

To simplify the calculation, natural channel is divided into three main subsections namely the right bank, the central reach and the left bank as shown in *Fig. 2.*

Equation (13) can be simplified into the following form:

$$K_{Ti} = k_l + k_c + k_r \tag{14}$$

where k_l, k_c and k_r are the conveyance of the left subsection, central channel, and right subsection respectively.

From the Manning formula, the conveyance of each subsection can be written as:

$$k_j = \frac{1}{nj} A_j R_j^{2/3} \qquad (j = l, c, r) \tag{15}$$

where l, c, and r denote the left subsection, central subsection and right subsection

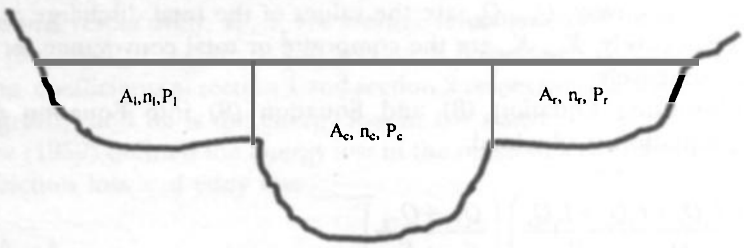


Fig. 2: Division of the flooded natural cross section into discrete elements

By substituting Equation (15) into Equation (14), the following equation is obtained:

$$h_f = \left(\frac{L_l Q_l + L_c Q_c + L_r Q_r}{Q_l + Q_c + Q_r}\right)\left(\frac{Q_{T1} + Q_{T2}}{K_{T1} + K_{T2}}\right) \tag{16}$$

where n_l, n_c, and n_r is Manning coefficient of roughness for left overbank, central channel, and right overbank respectively.

The velocity coefficient α at any river section can be written as:

$$WS_2 + \frac{\alpha_2 V_2^2}{2g} = WS_1 + \frac{\alpha_1 V_1^2}{2g} \tag{17}$$

where A_T is the total area of cross section, Al, Ac, Ar are flow area of left overbank, main channel and right overbank, respectively.

In Equation (17), the difference in velocity head between the main channel and the overbank sections is taken into consideration. The average velocity at a section can be described by:

$$V_i = \frac{Q_n}{A_n} \tag{18}$$

By substituting Equation (18), Equation (17) and Equation (11) into Equation (4) and after simplifying, the following equation is obtained:

$$WS_2 = WS_1 + \left(\frac{1-\theta}{2g}\right)\left[\left(\frac{Q_{T1}^2}{K_{T1}^3}\right)\left(\frac{k_{l1}^3}{A_{l1}^2} + \frac{k_{c1}^3}{A_{c1}^2} + \frac{k_{r1}^3}{A_{r1}^2}\right) - \left(\frac{Q_{T2}^2}{K_{T2}^3}\right)\left(\frac{k_{l2}^3}{A_{l2}^2} + \frac{k_{c2}^3}{A_{c2}^2} + \frac{k_{r2}^3}{A_{r2}^2}\right)\right]$$

$$+ \left(\frac{L_l\overline{Q}_l + L_c\overline{Q}_c + L_r\overline{Q}_r}{\overline{Q}_l + \overline{Q}_c + \overline{Q}_r}\right)\left(\frac{Q_{T1} + Q_{T2}}{K_{T1} + K_{T2}}\right)^2 \tag{19}$$

COMPUTATIONAL METHOD

Equation (19) describes the HEC-2 model that can be used to predict the water surface profile of a river for known values of discharge and the Manning coefficient of roughness. On the other hand the section geometry along the river must be defined in the model computation. Numerical computation of Equation (19) can be performed manually, but it is rather cumbersome and time consuming. The HEC-2 program has been developed to perform the numerical computation of Equation (19). The numerical implementation of Equation (19) can be explained as follows:

1. Assume a water surface elevation at the upstream cross section WS2 for subcritical flow in the river channel while the SW_1 is known.
2. Based on the assumed water surface elevation, determine the corresponding total conveyance. The determinations of the areas and the conveyance for subsections are important for model application.
3. Solve Equation (19) for SW_2 and compare the computed value of SW_2 with the value assumed in step 1; repeat steps 1 to 3 until the value agree within 0.01m accuracy. The calculated SW_2 will be used as SW_1 for the computation of the water surface elevation to the next upstream section.

Linggi River

The Linggi river is a major river system in the state of Negeri Sembilan, Malaysia. The river discharges its water to the sea (The Straits of Malacca) through the river mouth at Port Dickson, which is located approximately 53 km downstream from Seremban town, the capital of the state. Seremban is located approximately 70 km south of Kuala Lumpur at Latitude 2.750 North and Longitude 101.90 East. Simulation of water surface profile of the Linggi river was carried out for the river system in the vicinity of the Seremban town. *Fig. 3* shows the upstream Linggi River basin and the Seremban town occupies the lower portion of the basin. The Linggi river system passes through residential, commercial, industrial and agricultural areas. There are several areas within the township that experience flooding due to the high flows of the Linggi river during the rainy season. It is important to control flooding in the town centre by increasing the carrying capacity of the linggi river. According to a survey data, the average slope of the Linggi river is 1/500 while the measured length up to the control point (Station 2619401) is 5688 m (Mohammed 1998).

Data Acquisition

The data needed for this study was obtained from the Hydrology Section of the Drainage and Irrigation Department (DID) in Kuala Lumpur. The data acquired can be categorized as: (a) the streamflow records, (b) the stage records, (c) the longitudinal section of the river and (d) the cross sections of the river at 50 m interval.

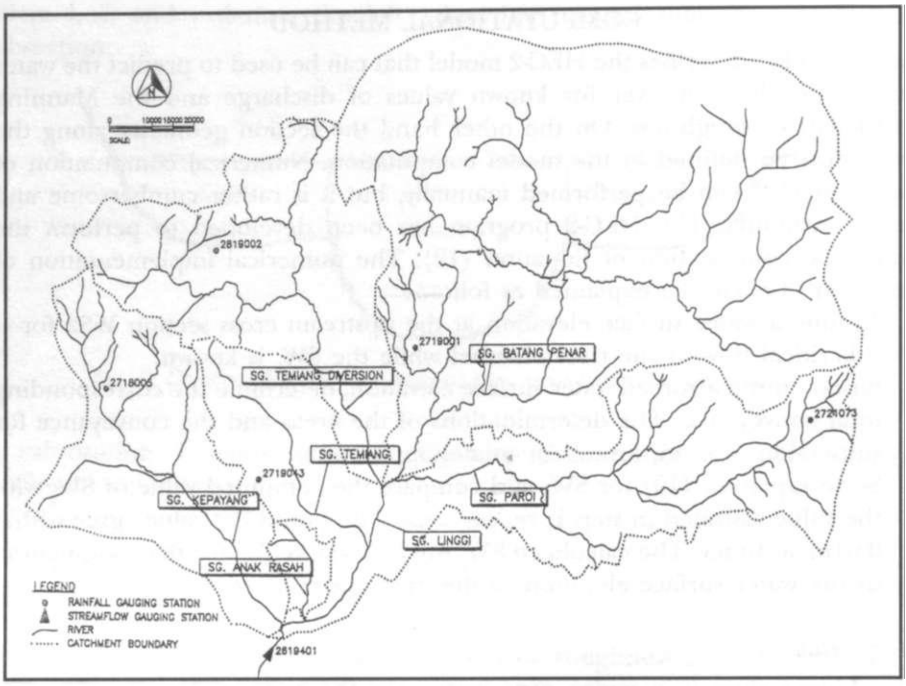


Fig. 3: Upstream basin of Linggi River system

Model Calibration and Verification

The calibration of HEC-2 model involves accurate estimation of the empirical hydraulic coefficients so that the flow events simulated by the model can produce flows of actual events as closely as possible. However, it is necessary to use the actual boundary conditions of the watercourse in the model. For backwater computation of the subcritical flow, the water level at a downstream control section is considered as a boundary condition in the HEC-2 model. This can be achieved by using the rating curve at this section. The other boundary condition that is involved is the tributary inflows to the main river. In the calibration process, consideration of various values of the incoming flow from tributaries to the main river will help the modeler to get accurate estimation for the roughness coefficients along the main watercourse. For the Linggi river, the eddy loss coefficient, θ and Manning coefficient of roughness, n were estimated based on field measurement of the water surface profile of the Linggi river for a stretch of 600 m using different discharges which include discharges of the tributaries. Applying the energy equation, values of the eddy loss coefficients for Linggi river were found to be within the range of values given by Chow (1959). The Manning formula was used to estimate the Manning coefficient of roughness, n along the stretch of the watercourse where the study was conducted, using various water levels.

To study the effect of the variation in values of Manning coefficient of roughness, the n values were varied from 0.03 to 0.032 for the central channel only. For the left and right banks which are grown with grass, the value of the Manning roughness coefficient were varied from 0.032 to 0.04 depending on the conditions of the banks. Field measurement for water surface profile for the Linggi river was used in the verification process of HEC-2 model. Using an actual flow of 36.2 m^3/s and a value of the Manning coefficient of roughness equal to 0.032, the HEC-2 model was used to predict the water surface profile for the 600 m stretch of the Linggi river. The simulated water surface profile was then compared to the actual measured water surface profile as shown in *Fig. 4*. The maximum and minimum absolute errors in the predicted water surface profile for Linggi river were computed and found to be 100 mm and 20 mm respectively (Mohammed 1998). The deviations were less than 5%.

SENSITIVITY ANALYSIS

Accuracy of the computed water surface profiles for a river using HEC-2 model is dependent on many factors such as the accuracy of the stream geometry, accurate estimation of the Manning coefficient of roughness and interval between stations along a river. It is premised that modern survey technologies can give good accuracy in determining cross-sectional geometry of the river, so the sensitivity analysis was limited to the impact of Manning coefficient of roughness and the intervals between the cross sections or stations on the accuracy of the predicted water surface profile for Linggi river using HEC-2 model. Using constant interval of 50 m between stations along the Linggi river,

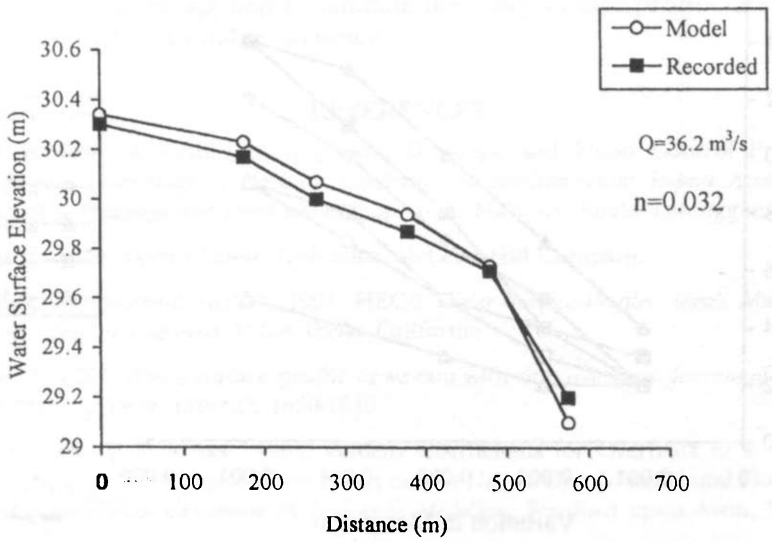


Fig. 4: Verification of the HEC-2 model using recorded water surface elevation of the Linggi river

a variation of 0.001 in the value of Manning coefficient of roughness was associated with an average variation of 2.5 cm in the prediction water surface elevation (Mohammed 1998). *Fig. 5* shows the impact of the variation in Manning coefficient of roughness on the predicted water surface elevation at 1 km interval along the stretch of the Linggi river. A difference of 51 cm, 37 cm, 7 cm, 3 cm in the predicted water surface elevation at the most upstream section were obtained by running the HEC-2 computer package for a stretch of 5688 m from Linggi river using intervals of 1000 m, 500 m, 200 m, and 100 m receptively. These differences were computed using the predicted water surface elevation for 50 m interval. The bigger errors in the longer interval predictions may be attributed to the accumulated errors in misrepresenting the real features of the stream geometry and alignment.

Model Application

The model was used to predict the water surface profile of the Linggi river for flood occurrences of Q_2, Q_{10}, and Q_{100}. The values obtained were 100 m^3/s for 100-year return period, 57.2 m^3/s for 10-year return period and 32.7 m^3/s for 2-year return period. Cross sections at 50 m interval along the watercourse of the Linggi River for a reach of 5.688 km were used as the input data to the HEC-2 model. The value of Manning coefficient of roughness being used for central channel was 0.030 while a value of 0.035 was used for both the right overbank and left overbank. *Fig. 6* shows the predicted water surface profiles for Linggi river for different flood magnitudes (Mohammed 1998).

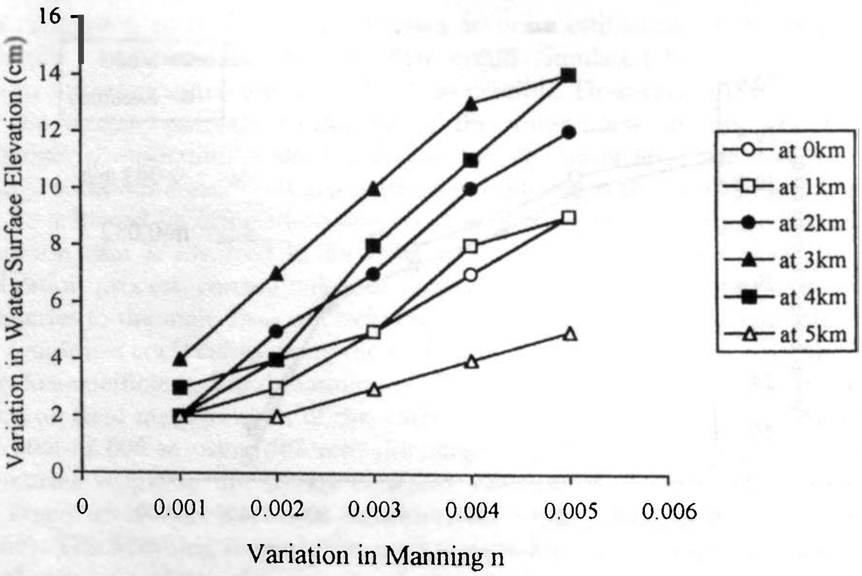


Fig. 5: Effect of Manning roughness on the predicted water surface elevation along the Linggi River

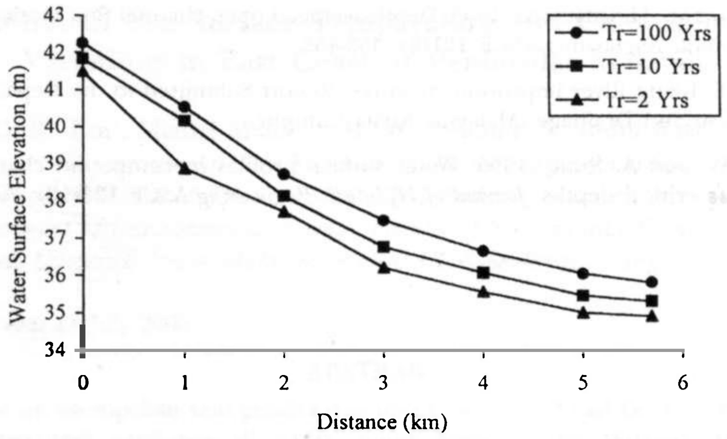


Fig. 6: Predicted water surface elevation for the Linggi River

CONCLUSION

The application of the HEC-2 model to the Linggi river system in the vicinity of Seremban, showed a good agreement between the predicted water surface profile and the recorded water surface profile. The maximum absolute error in the predicted water surface profile for Linggi river was found to be 100 mm while the minimum absolute error was only 20 mm only. These errors are less than 5% of the recorded data). Model testing showed that HEC-2 is sensitive to value of Manning coefficient of roughness and the intervals of cross sections long the studied river system. Tacking into consideration model sensitivity, HEC-2 model can be applied to simulate the water surface profile of a tropical river system with reasonable accuracy.

REFERENCES

ANN, O.C. 1994. A Review of Irrigation, Drainage and Flood Control Projects in Malaysia. *Proceedings of National Conference on Environmental Impact Assessment for Irrigation Drainage and Flood Control*, pp. 1-14. Malaysia: Kuala Terengganu.

CHOW, V.T. 1959. *Open Channel Hydraulics.* McGraw-Hill Company.

HYDROLOGIC ENGINEERING CENTER. 1991. HEC-2 *Water Surface Profiles: User's Manual U.S. Army Corps of Engineers,* U.S.A Davis: California.

ISHIKAWA, T. 1984. Water surface profile of stream with side overflow. *Journal of Hydraulic Engineering* ASCE **110(12)**: 1830-1840.

LYNESS, J.F. and W.R. MYERS. 1994. Velocity Coefficients for Overbank in a Compact Compound Channel and Their Effect on the Use of One Dimensional Flow Models *2nd International Conference on Hydraulic Modelling,* Stanford upon Avon, U.K.

MOHAMMED, THAMER AHMED. 1998. Flood simulation models for a river system in a tropical region: the case of Linggi River Malaysia. A Ph.D Thesis, Faculty of Engineering, Universiti Putra Malaysia.

MOLLS, T. and M. H. CHAUDHRY. 1995. Depth-averaged open-channel flow model. *Journal of Hydraulic Engineering* ASCE **121(6)**: 453-465.

NIK, A. 1996. Klang River improvement works. Report Submitted to the Department of Irrigation and Drainage. Malaysia: Kuala Lumpur.

STURM, T. W. and A. SADIQ. 1996. Water surface profiles in compound channel with multiple critical depths. *Journal of Hydraulic Engineering* ASCE **122(12)**: 703-708.

ISSN: 0128-7680
Pertanika J. Sci. & Technol. 10(1): 13-24 (2002)

Studies of Sea Surface Temperature and Chlorophyll-a Variations in East Coast of Peninsular Malaysia

C. K. Tan[1], Shattri Mansor[1], H. M. Ibrahim[2] & Abdul Rashid[1]
[1]*Department of Civil Engineering, Faculty of Engineering Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*
[2]*Department of Environmental Science, Faculty of Sciences and Environmental Studies, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*

Received: 13 July 2000

ABSTRAK

Kajian ini merupakan satu pendekatan untuk mengklasifikasi keadaan laut di kawasan grid perikanan di pantai timur Semenanjung Malaysia dengan menggunakan data klorofil -a dan suhu permukaan laut dari satelit. Perubahan suhu permukaan laut dan klorofil-a di Laut China Selatan adalah bergantung kepada sistem monsun. Keputusan daripada analisis ini menunjukkan perubahan suhu permukaan laut dan klorofil-a turut bergantung pada lokasi grid-grid perikanan ini. Pengkelasan bagi klorofil-a menunjukkan perbezaan yang jelas berbanding suhu permukaan laut. Klasifikasi hierarki dapat membantu dalam memahami perubahan keadaan laut di antara grid-grid perikanan. Walau bagaimanapun, untuk lebih memahami kesan perubahan keadaan laut ke atas sumber marin di EEZ Malaysia, kajian yang lebih mendalam perlu dijalankan dengan mengguna data jangka panjang.

ABSTRACT

This paper attempts to classify the oceanographic conditions of the fishing grids in east coast of Peninsular Malaysia using surface chlorophyll-a content and sea surface temperature (SST) data from satellite. The variation of SST and chlorophyll-a content in the South China Sea is greatly affected by the monsoon system. Analysis results showed that both SST and chlorophyll-a variations of the fishing grids are closely related to their geographical locations. The classification using chlorophyll-a on the fishing grids give a clearer variation compared to SST. Hierarchical cluster analysis gave a better means of understanding the variations of these oceanographic conditions and the relationship among the fishing grids. However, to understand how these variations of oceanographic condition affect the marine fisheries catch in Malaysian Exclusive Economic Zone (EEZ), further studies should be conducted using longer time scale data.

Keywords: South China Sea, remote sensing, geographical information system (GIS), sea surface temperature, chlorophyll-a, classification, fishing grid

INTRODUCTION

The fisheries sector has played an important role in the Malaysian economy with regards to foreign exchange earnings, employment opportunity, income and supply of protein to the populace. Basically, the fisheries industry in Malaysia can be divided into marine fisheries, aquaculture and inland fisheries.

With the declaration of Malaysia's Exclusive Economic Zone (EEZ) in 1980, the coverage of Malaysia's waters expanded from 47,000 n.m.2 to 162,000 n.m.2. This increase of maritime area under national jurisdiction provided greater potential to explore the offshore marine resources especially in the South China Sea.

The South China Sea is generally shallow (50m), and sitting largely on the vast Sunda Shelf platform. The climate of this area is influenced monsoons, i.e. the northeast monsoon (November to February) and the southwest monsoon (May to September). Two transitional periods can be clearly distinguished between these two seasonal monsoon winds (Nasir *et al.* 1999).

The South China Sea demonstrates the characteristics of tropical waters. Generally, the sea surface temperature is relatively constant. According to Mohsin and Ambak (1996), SST range from 28 – 32°C in warmer months (May/June), while in cooler months (December/January) it ranges from 25.2 – 28.8°C.

The chlorophyll-a content in the east coast of Peninsular Malaysia is generally low, ranging from 0.08 (surface) to 0.36 mg/m^3 (50m) off Terengganu waters and 0.0269 (surface) to 0.2434 mg/m^3 off Pahang waters (Ahmad & Ichikawa 1986; Rahman 1986; Mohsin and Ambak 1996). Normally the highest chlorophyll concentration is located at the chlorophyll maximum layer, at a depth of 25-50m.

There has been little written about the variations of physical and biological parameters and their relationships in this region. This could be due to little opportunity to study the phenomena on a wider spatial scale. Satellite remote sensing technology has created this opportunity to study the oceanography on a wider scale by providing a synoptic view and observation of the oceanographic parameters. This paper is aimed at studying the sea surface temperature (SST) and chlorophyll-a seasonal variations by using satellite remote sensing, and classifies the fishing grids based on these two parameters.

MATERIALS AND METHODS

The area of study covers only the EEZ of the east coast of Peninsular Malaysia. This area has been divided into 50 fishing grids, which cover the whole EEZ with the total of 45000n.m.^{2}. The size area of each grid cell is 30 n.m.x 30 n.m.(total 900n.m.^{2}). The fishing grids in the east coast of Peninsular Malaysia, which is used for catch log reporting, is shown in *Fig. 1*.

Monthly sea surface temperature and chlorophyll-a data were acquired from National Space Development Agency of Japan (NASDA), using the Advance Earth Observing Satellite (ADEOS)/Ocean Colour and Temperature Scanner (OCTS) Global Map Data Set. The study period was from November 1996 to June 1997.

The satellite images were processed then imported into ArcView GIS 3.1 with Spatial Analyst. Images were resampled into the size of the fishing grids (0.5° x 0.5°). The SST and chlorophyll-a values were then inputted into SPSS

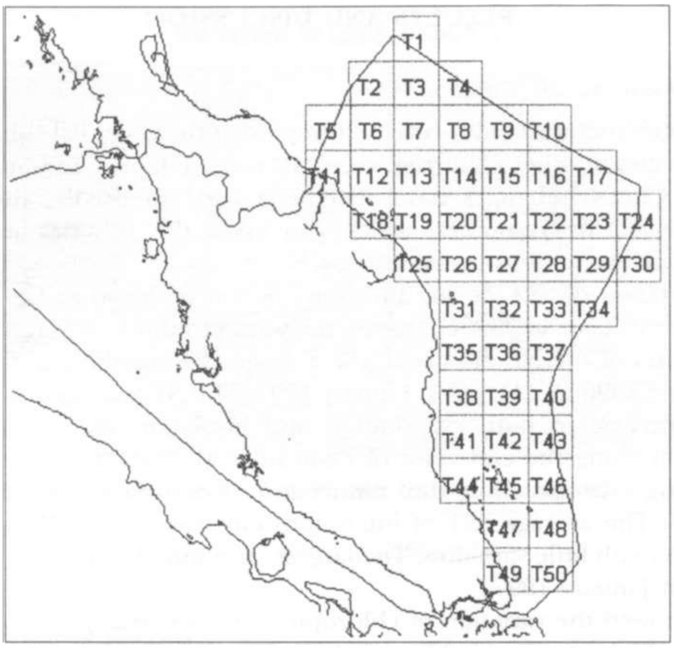


Fig. 1: Fishing grids in the east coast of Peninsular Malaysia

6.0 for the hierarchical clustering analysis (Ward method). Dendrograms were produced as the result of the clustering analysis. Results of statistical analysis were then imported into ArcView again to produce classification maps for visualization purpose. The flow chart of the methodology is shown in *Fig. 2.*

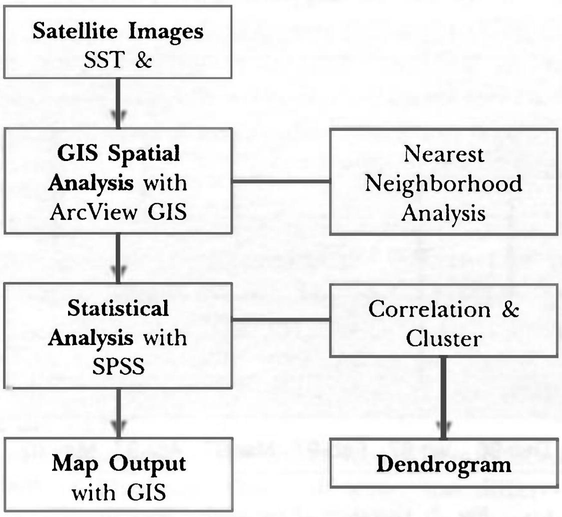


Fig. 2: Flow chart of methodology

RESULTS AND DISCUSSION

Temporal Oceanographic Variation

Sea surface temperature has been used to predict the potential fishing grounds of some species in other countries, such as America, Japan, Taiwan, etc (Rao *et al.* 1999). Climate changes have also been used to predict the long-term changes of marine resources, which can assist the fisheries institution to manage their marine resources, effectively.

The variation of SST during the study period is shown as *Fig. 3.* The SST ranged from 25.65°C to 33.9°C. During the warmer month of May to September with the onset of the NE monsoon, the average SST significantly from 32.29°C in November 1996 to 29.03°C in January 1997. The NE monsoon season causes a major decrease of both evaporation and insolation values and increased precipitation along the east coast of Peninsular Malaysia. Overcast skies block the incoming solar radiation, thus reducing the sea surface temperature (Nasir *et al.* 1999). The average SST of inter-monsoon season and SW monsoon was around 31°C with little variation. The largest variation of SST among the fishing grids was in January 1997.

Fig. 4 showed the variation of chlorophyll-a concentration during the study period. The highest average chlorophyll-a value of the study area was 0.585mg/m^3 in December 1996 at the beginning of the NE monsoon season. Chlorophyll-a concentrations were low during the inter-monsoon season (March - April) until May 1997, and then increased significantly to 0.271 mg/m^3 in June1997. Chlorophyll-a distribution has larger variations among the fishing grids compared to SST.

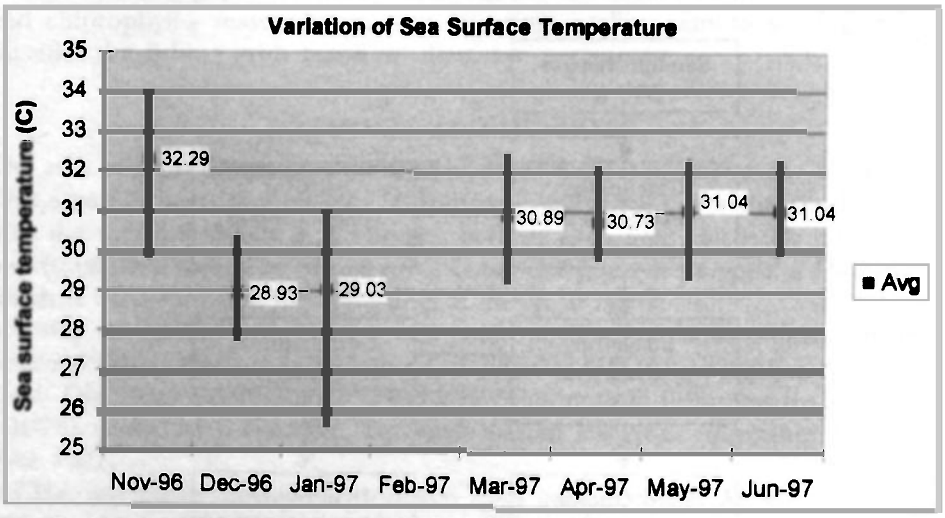


Fig. 3: Variation of sea surface temperature

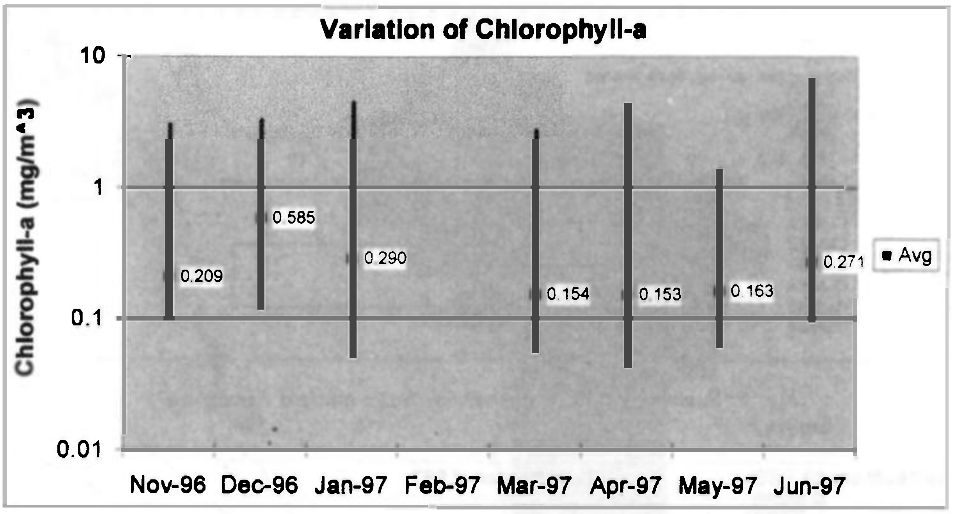


Fig. 4: Variation of chlorophyll-a

Cluster Classification Analysis

Sea Surface Temperature

The summary of SST dendrogram is shown in *Fig. 5*. From this result, SST in the fishing grids can be classified into 6 classes. The average SST of each class is displayed in *Fig. 6*. Class 1 and 2 demonstrate some similarity in their trends, where their average SST decreased during April 1997. Besides, class 5 and 6 also show similar trend except in May 1997, where class 5 decreased while class 6 kept on increasing.

Fig. 7 shows the result of correlation analysis among the classes. According to the correlation between the classes, class 5 and 6 has the highest R2 value 0.984, followed by class 1 and 6 with R2 equals to 0.957. Although class 1 and 2 has been classified as a group by using cluster analysis, their R^2 only noted at 0.95. Classes 1 and 5 also have a good correlation, with R^2 equals to 0.944. The R^2 between class 3 and 4 is 0.903. Classes 2 and 3 have the lowest correlation with R^2 equals to 0.717.

The SST classification results were produced on a map shown in *Fig. 8*. Classes 5 and 6 are both closely located at the eastern part of the EEZ. The grids of class 6 seem to border class 5. The SST trend and variation of both classes are different from others. This may due to their topology where all the grids are located at the off shore areas. Besides, they also have a good correlation with their neighboring grids, class 1, which is located at the northern part of the EEZ.

Class 1 has been classified to the other major group because it is closely located to the Gulf of Thailand. The Gulf water has different characteristic from the open sea, it may give a significant effects to the grids in class 1. The Gulf water intrusion phenomenon was first reported during the Matahari

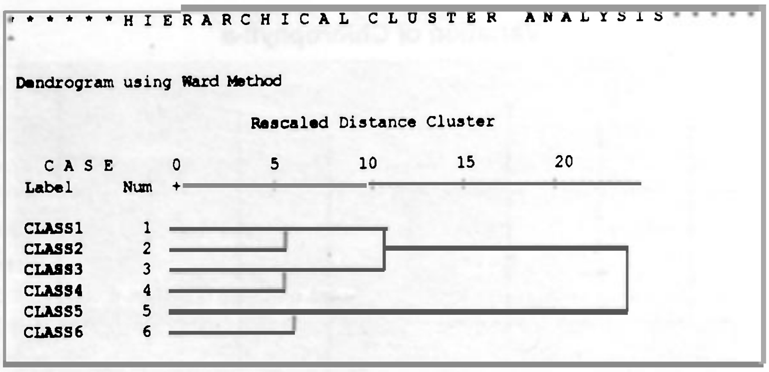


Fig. 5: Summary of SST hierarchical cluster analysis dendogram

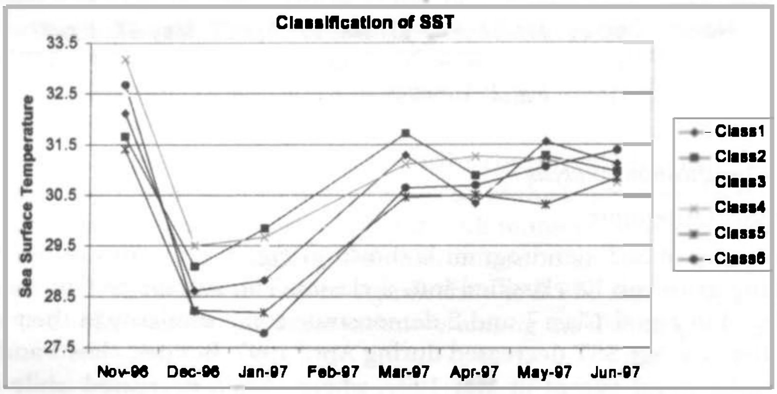


Fig. 6: Variation of average SST for each class

- - Correlation Coefficients - -

	CLASS1	CLASS2	CLASS3	CLASS4	CLASS5	CLASS6
CLASS1	1.0000 (7) P= .	.9502 (7) P= .001	.8665 (7) P= .012	.8747 (7) P= .010	.9444 (7) P= .001	.9569 (7) P= .001
CLASS2	.9502 (7) P= .001	1.0000 (7) P= .	.7167 (7) P= .070	.8251 (7) P= .022	.9092 (7) P= .005	.8746 (7) P= .010
CLASS3	.8665 (7) P= .012	.7167 (7) P= .070	1.0000 (7) P= .	.9028 (7) P= .005	.8590 (7) P= .013	.9272 (7) P= .003
CLASS4	.8747 (7) P= .010	.8251 (7) P= .022	.9028 (7) P= .005	1.0000 (7) P= .	.8789 (7) P= .009	.9136 (7) P= .004
CLASS5	.9444 (7) P= .001	.9092 (7) P= .005	.8590 (7) P= .013	.8789 (7) P= .009	1.0000 (7) P= .	.9840 (7) P= .000
CLASS6	.9569 (7) P= .001	.8746 (7) P= .010	.9272 (7) P= .003	.9136 (7) P= .004	.9840 (7) P= .000	1.0000 (7) P= .

Fig. 7: Correlation analysis among the classes

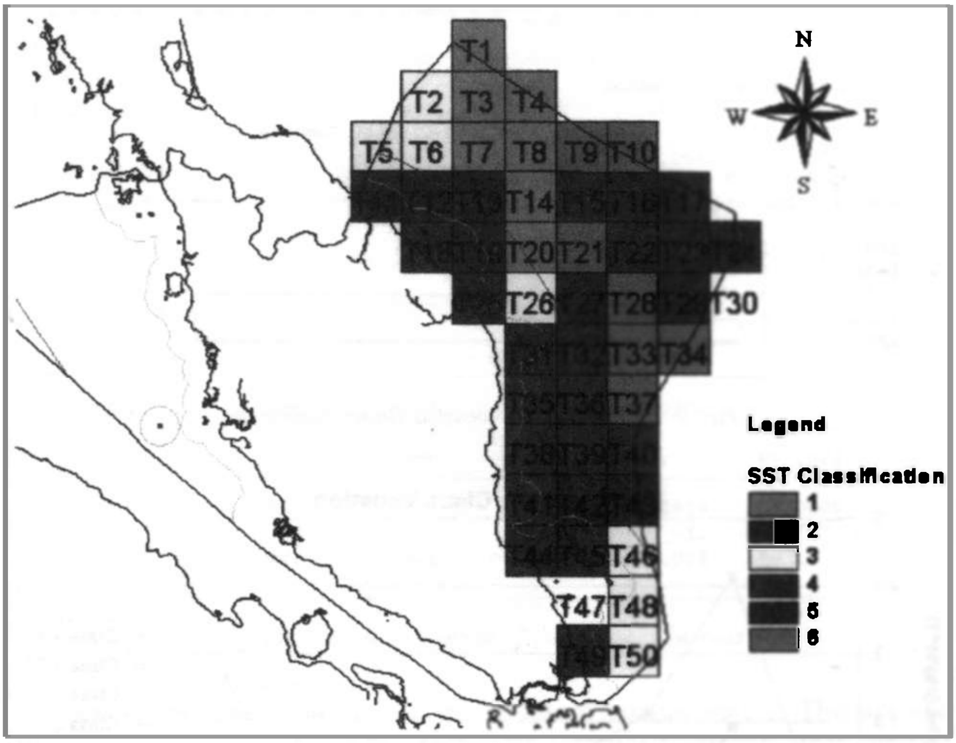


Fig. 8: Map of SST classification results

Expedition' 85 by Kawamura (1986) in May 1985. The NOAA satellite image shown that there are two different types of surface water off Terengganu, where the Gulf water coming down southwards and offshore waters going up northwards (Kawamura 1986).

Class 4 grids generally located at the southern areas. According to Nasir *et al.* (1999), the southern tip of the east coast of Peninsular Malaysia shows a different variation of physical properties (temperature, salinity and density profile) compared to the northern part.

Chlorophyll-a

The summary of chlorophyll-a dendrogram is shown in *Fig. 9.* From the result, all the chlorophyll-a variations in the fishing grids can be classified into three major classes. *Fig. 10* illustrates the average chlorophyll-a variation for each class. Class 1 is further classified into three subclasses, namely class 1.1, class 1.2 and class 1.3. All the chlorophyll-a classes reached their peak value in December 1996 during the NE monsoon.

Class 1.2 is generally located at the northern part of the EEZ. It is greatly affected by the characteristics of Gulf water. Class 1.2 also showed highest chlorophyll-a values (about 1.5 mg/m^3) among the subclasses of class 1 during

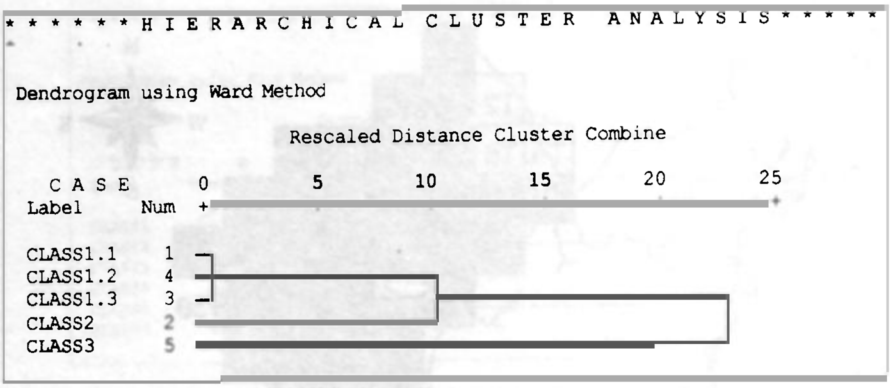


Fig. 9: Summary of chlorophyll-a cluster dendogram

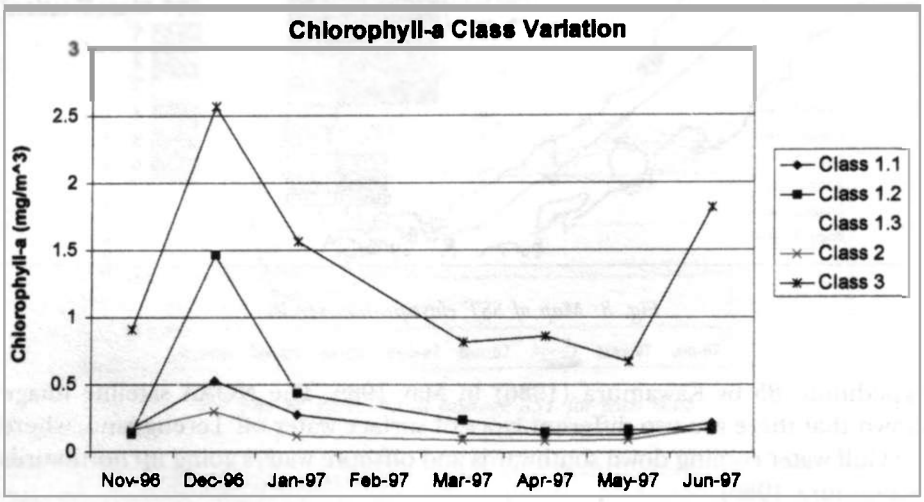


Fig. 10: Chlorophyll-a variation for each class

December 1996. According to Sopana (1999), the northeast monsoon was well developed in December. The upwelling occurred at the east coast of the Gulf of Thailand and the Vietnamese coast in January. During this time, the surface current flowed from the Vietnamese coast to the lower Gulf and east coast of Peninsular Malaysia until March. Thus, the rich Gulf water may cause the increment of chlorophyll-a at class 1.2 waters during the NE monsoon.

Class 1.1 is located close to class 1.2 and some at the southern part of the EEZ. It showed high correlation with class 1.2 and class 1.3, where the R^2 value equals to 0.964 and 0.945 respectively, which is the highest among the classes. Class 1.1 seems to be the intermediate type of waters between class 1.2 at the northern part and class 1.3 at the southern part, where its average chlorophyll-a variation in *Fig. 10* also between these two classes. *Fig. 11* shows the correlation analysis among the classes.

- - Correlation Coefficients - -

	CLASS1.1	CLASS1.2	CLASS1.3	CLASS2	CLASS3
CLASS1.1	1.0000	.9642	.9448	.8989	.9301
	(7)	(7)	(7)	(7)	(7)
	P= .	P= .000	P= .001	P= .006	P= .002
CLASS1.2	.9642	1.0000	.8692	.8300	.8287
	(7)	(7)	(7)	(7)	(7)
	P= .000	P= .	P= .011	P= .021	P= .021
CLASS1.3	.9448	.8692	1.0000	.8040	.9310
	(7)	(7)	(7)	(7)	(7)
	P= .001	P= .011	P= .	P= .029	P= .002
CLASS2	.8989	.8300	.8040	1.0000	.8879
	(7)	(7)	(7)	(7)	(7)
	P= .006	P= .021	P= .029	P= .	P= .008
CLASS3	.9301	.8287	.9310	.8879	1.0000
	(7)	(7)	(7)	(7)	(7)
	P= .002	P= .021	P= .002	P= .008	P= .

Fig. 11: Correlation analysis of the chlorophyll-a class

Class 1.3 is mainly located in the southern coastal waters. The average chlorophyll-a value of class 1.3 is higher than class 1.1 (*Fig.10*). According to Liong (1974), upwelling during the SW monsoon at the southern part of east coast of Peninsular Malaysia results in the higher primary productivity of the surface water during the period. Thus, cause the increment of class 1.3 chlorophyll-a contents in April until June 1997.

Class 2 is located at the offshore areas; their surface chlorophyll-a content is the lowest compared to the others. This result agreed with the work of other researchers, which showed the offshore water in South China Sea is generally low in chlorophyll-a. Low chlorophyll-a values at the surface could be attributed to low nutrient levels compare to deeper layers, as well as availability of too much irradiant energy (Shamsudin & Awang 1993).

Class 3 is located close to the river mouth and all are located at the coastal area. Their chlorophyll-a content is the highest among the classes. This may due to the inflow of water from the river that carries a lot of land-based nutrients that enrich the coastal water and increase their primary productivity. The map of the chlorophyll-a classification result is shown in *Fig. 12*.

CONCLUSION

The variations of sea surface temperature and chlorophyll-a content in the South China Sea is greatly affected by the monsoon system. The northeast monsoon season causes a major decrease in SST values along the east coast of Peninsular Malaysia. Sea surface temperature of the water dropped significantly

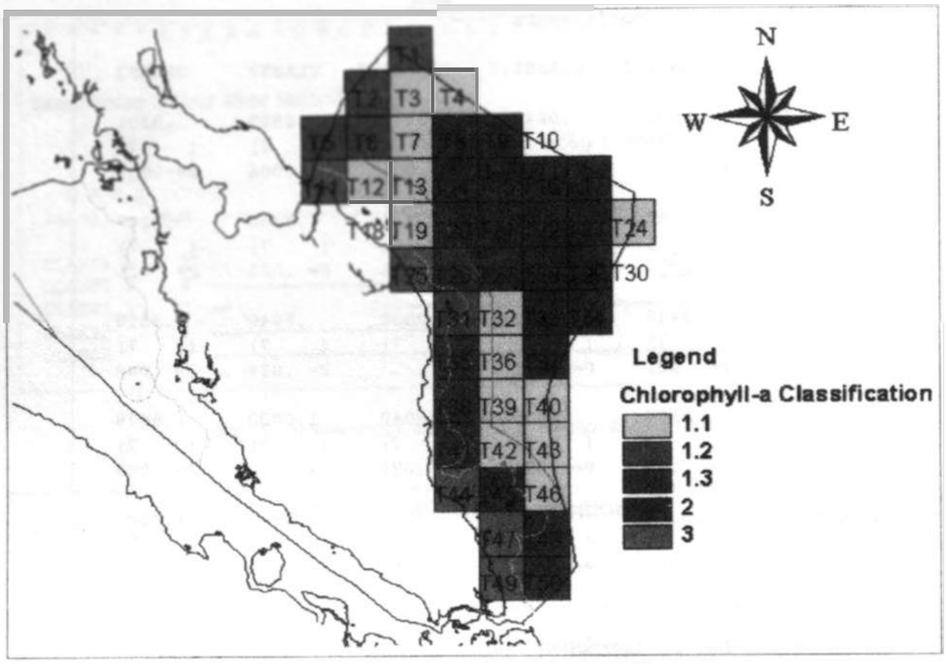


Fig. 12: Map of the chlorophyll-a classification results

during December 1996. Conversely, chlorophyll-a value reached the peak of the study period in December 1996.

During the inter-monsoon season, the chlorophyll-a content was low, while the SST values almost remain constant at about 31°C in April. Only when the SW monsoon started in May 1997, the chlorophyll-a content increased significantly, while the SST remained the same.

The classification of SST and chlorophyll-a variation of the fishing grids was successful, where the grids were classified into 6 major classes using SST, and another 3 major classes using chlorophyll-a. The results show that both SST and chlorophyll-a variation of the fishing grids are closely related to their geographical location.

Sea surface temperature variation of the offshore grids seems to be different from other classes. For the grids that are located at the northern part of the EEZ, the SST variations may be affected by the Gulf of Thailand waters.

The classification using chlorophyll-a on the fishing grids give a clearer classification compared to SST. The coastal water and offshore water showed major differences between them. Chlorophyll-a content is low among the offshore waters (average 0.148 mg/m^3), while the coastal waters, close to the river mouths, seem to have a very high chlorophyll-a content (average 1.315mg/m^3).

This paper attempts to classify the oceanographic conditions of the fishing grids in the east coast of Peninsular Malaysia using surface chlorophyll-a content and sea surface temperature data from satellite. This is a first step towards determination of potential fishing zones for fishery management. The hierarchical cluster analysis gave a better means of understanding to the variation of these oceanographic conditions and the relationship among the fishing grids. However, to understand how these variations of oceanographic condition affect the marine fisheries catch in Malaysian EEZ, further studies should be conducted employing longer time scale data.

ACKNOWLEGDEMENT

The authors would like to thank Mr. Hiroyuki Yanagawa, JICA expert on fishery resources, for helping us in the interpretation of cluster analysis and for provided valuable suggestions on preparing this paper. This project on fish forecasting by satellite is supported by Intensification of Research in Priority Areas (IRPA) Funds hereby acknowledge.

REFERENCES

AHMAD, RAIHAN B. SHEIKH and TOSHIHIRO ICHIKAWA. 1986. Chlorophyll-a content off the Terengganu Coast. In *Ekspedisi Matahari' 85: A Study on the Offshore Waters of the Malaysian EEZ*, ed. Abu Khair Mohammad Mohsin, Mohd. Ibrahim Hj. Mohamed & Mohd. Azmi Ambak, p. 117-120. Faculty of Fisheries and Marine Science, Universiti Pertanian Malaysia, Serdang, Selangor. Occasional Publication No.3.

KAWAMURA, GUNZO. 1986. A preliminary survey of the sea surface in Malaysian waters and the Gulf of Thailand with the aid of a satellite data utilization system. In *Ekspedisi Matahari' 85: A Study on the Offshore Waters of the Malaysian EEZ*, ed. Abu Khair Mohammad Mohsin, Mohd. Ibrahim Hj. Mohamed & Mohd. Azmi Ambak, p. 45-48. Faculty of Fisheries and Marine Science, Universiti Pertanian Malaysia, Serdang, Selangor. Occasional Publication No.3.

LIONG, PIT CHONG. 1974. Hydrographic conditions off the east coast of Peninsular Malaysia during August 1972. *Malaysian Agriculture Journal* **49**: 492-513.

MOHSIN, ABU KHAIR MOHAMMAD and MOHD. AZMI AMBAK. 1996. *Marine Fishes & Fisheries of Malaysia and Neighbouring Countries.* Serdang: Universiti Pertanian Malaysia Press.

NASIR, M.S., PENJAN ROJANA-ANAWAT and ANOND SNIDVONGS. 1999. Physical characteristics of watermass in the South China Sea, Area I: Gulf of Thailand and East Coast of Peninsular Malaysia. In *Proceedings of The First Technical Seminar on Marine Fishery Resources Survey in the South China Sea, Area I: Gulf of Thailand and Peninsular Malaysia,* p. 1-5. 24-26 November 1997. Samutprakan, Bangkok: SEAFDEC.

RAHMAN, RIDZWAN BIN ABDUL. 1986. Chlorophyll "a" content of the south-western portion of the South China Sea. In *Ekspedisi Matahari' 86: A Study on the Offshore Waters of the Malaysian EEZ*, ed. Abu Khair Mohammad Mohsin, Ridzwan Abdul Rahman & Mohd. Azmi Ambak, p. 61-63. Faculty of Fisheries and Marine Science, Universiti Pertanian Malaysia. Occasional Publication No.4.

RAO, K.H., M.V. RAO, S.B. CHOUDHURY, S.K. SASAMAL and A.N. NATH. 1999. Potential Fishing Zone Mapping: An over view. *Geoinformatics: Beyond 2000.* 9-11 March 1999. p. 352-372. Dehra Dun, India.

SHAMSUDIN, LOKMAN & ABDUL AZIZ AWANG. 1993. Biomass chlorophyll "a" values of the coastal waters around Peninsular Malaysia. Preliminary report of the KL Paus Oceanographic Research Project, 1993.

SOPANA, BOONYAPIWAT. 1999. Distribution, abundance and species composition of phytoplankton in the South China Sea, Area I: Gulf of Thailand and east coast of Peninsular Malaysia. In *Proceedings of the First Technical Seminar on Marine Fishery Resources Survey in the South China Sea, Area I: Gulf of Thailand and Peninsular Malaysia*, p. 111-134. 24-26 November 1997, Samutprakan, Bangkok: SEAFDEC.

ISSN: 0128-7680
Pertanika J. Sci. & Technol. **10(1)**: 25-33 (2002)

The Performance of AICC as an Order Selection Criterion in ARMA Time Series Models

Liew Khim Sen & Mahendran Shitan[a]
Department of Mathematics
Universiti Putra Malaysia
43400 UPM, Serdang
Selangor, Malaysia

Received: 4 December 2000

ABSTRAK

Kajian ini bertujuan untuk menilai prestasi kriteria maklumat Akaike diperbaiki atau AICC (Akaike's Information Corrected Criterion) sebagai kriteria penentuan peringkat dalam pembentukan model Autoregresif Purata Bergerak (Autoregressive Moving-Average) atau ARMA $(p.q)$. Suatu penyelidikan simulasi dijalankan untuk menentukan kebarangkalian kriteria AICC minimum telah memilih model *sebenar* dengan tepat. Keputusan yang diperolehi menunjukkan bahawa prestasi AICC adalah sekadar sederhana. Masalah lebihan pembolehubah (over pemeterization) berada pada tahap yang minimum. Oleh itu, bagi sebarang dua model yang setanding, adalah lebih wajar untuk memilih model dengan peringkat p dan q yang lebih rendah.

ABSTRACT

This study is undertaken with the objective of investigating the performance of Akaike's Information Corrected Criterion (AICC) as an order determination criterion for the selection of Autoregressive Moving-Average or ARMA (p,q) time series model. A simulation investigation was carried to determine the probability of the AICC statistics picking up the *true* model. Result obtained showed that the probability of the AICC criterion picking up the correct model was moderately good. The problem of over parameterization existed but under parameterization was found to be minimal. Hence, for any two comparable models, it is always safe to choose the one with lower order of p and q.

Keywords: AICC, ARMA, under/over parameterization

INTRODUCTION

In the process of time series autoregressive moving-average or ARMA (p, q) modelling, we do not know the *true* order of the model generating the data. In fact it will usually be the case that there is no true ARMA (p, q) model, in which case our goal is simply to find one that represents the data optimally in some sense (Brockwell and Davis 1996). However, the challenge is to decide the optimal orders of p and q (Beveridge and Oickle 1994). In a given application,

Note: We are indebted to two anonymous referees for their helpful comments and suggestions on a previous draft of this paper. All errors remain our responsibility the Box-Jenkins model selection procedure may suggest several specifications, each of which satisfies the diagnostic checks. Some kind of a measure of goodness of fit is therefore needed to distinguish between different models in these circumstances (Harvey 1993). Many criteria have been suggested for this reason by the past researchers. The Akaike's information corrected criterion (Hurvish and Tsai 1989) or AICC, among others, is a commonly used criterion. However, its performance must be evaluated. Therefore, the objective of the study is to evaluate the performance of AICC statistics in selecting the true ARMA time series model based on a simulation study.

The rest of this paper is organised as follows. The next section discussed the order determination criterion. This is followed by a description of simulation study and a report of simulation result. Finally, the conclusions of the study are presented.

ORDER DETERMINATION CRITERIA

Many criteria has been proposed for the purpose order determination by past researcher. These include the final prediction error (EPE) criterion, Schwarz-Rissanen criterion (SIC), Bayesian estimation criterion (BEC), Hannan- Quin criterion, Akaike's information criterion (AIC) and so on. The latest model selection criterion is the Akaike's information corrected criterion AICC, developed by Hurvish and Tsai in 1989.

There has been considerable literature published on order determination criteria. A brief discussion of these criteria is available in Beveridge and Oikle (1994); de Gooijer *et al.* (1985) and Stoica *et al.* (1986). Brockwell and Davis (1996) present greater theoretical and practical detail and additional references for many of these criteria.

The final prediction error, RPE criterion was original proposed by Akaike (1969, 1970) for AR (p) order determination and was extended to ARMA (p,q) models by Sōderstrō, in 1977 (Beveridge and Oickle 1994). This criterion was established on the basis of minimizing the one-step-ahead mean square forecast error after incorporating the inflating effects of estimated coefficients. The criterion to be minimized is

$$\text{FPE} = \hat{\sigma}^2 \frac{n+p+q}{n-p-q} \tag{1}$$

Where $\hat{\sigma}^2$ is estimated variance of white noise,
n is number of observation,
p is order of the autoregressive component,
and q is order of the moving average component.

In 1970, Akaike found that FPE is asymptotically inconsistent and in 1973 he employed information-theoretic considerations to develop the Akaike's information criterion, AIC. This was designed to be an asymptotically unbiased estimate of the Kullback-Leibler index of the fitted model relative to the true model (Akaike 1973). The AIC statistics is defined as

$$\text{AIC} = -2 \ln \text{Likelihood } (\hat{\phi}, \hat{\theta}, \hat{\sigma}^2) + 2(p + q + 1) \quad \textbf{(2)}$$

where $\hat{\phi}$ are estimated autoregressive parameters,
$\hat{\theta}$ are estimated moving average parameters,
and $\hat{\sigma}^2$, n, p and q are as defined in equation (1).

A criterion like AIC that penalizes the likelihood for the number of parameters in the model attempts to choose the most parsimonous model. However, AIC is only asymptotically unbiased and Jones (1975) and Shibata (1976) showed empirical evidence that AIC has the tendency to pick models which are over-parameterized. In view of this, Akaike applied a Bayesian modification to AIC and finally in 1978, he came up with a consistent order selection criterion, known as Bayesian information criterion or BIC (see Akaike 1979). If the data $\{X_1, ..., X_n\}$ are in fact observations of an ARMA (p, q) process, then a Bayesian information criterion is defined to be

$$\text{BIC} = (n - p - q)\ln \frac{n\hat{\sigma}^2}{n-p-q} + n(1 + \ln 2n) + (p+q)\ln\left[\frac{\sum_{t=1}^{n} X_t^2 - n\hat{\sigma}^2}{p+q}\right] \quad (3)$$

There is evidence to suggest that the BIC is more satisfactory than the AIC as an ARMA model selection criterion since the AIC has a tendency to the pick models, which are over-parameterized (Hannan 1980).

Schwarz (1978) used a Bayesian analysis and Rissanen (1978) applied an optimal data-recording scheme to independently arrive at the same criterion, later known as Schwarz-Rissanen criterion, SIC. The criterion to be minimized is given by

$$\text{SIC} = \ln \hat{\sigma}^2 + \left(\frac{p+q}{n}\right) \ln n \quad \textbf{(4)}$$

Geweke and Mease (1981) suggested approximating SIC by Bayesian estimation criterion, BEC.

$$\text{BEC} = \hat{\sigma}^2 + (p_x + q_x)\,\hat{\sigma}^2_x \ln \frac{n}{n - p_x - q_x} \quad \textbf{(5)}$$

where x denotes a quantity from pre-assigned high order ARMA model that includes all potential models.

Hannan and Quinn (1979) and Hannan (1980) constructed Hannan-Quinn criterion from the law of the iterated logarithm. It provides a penalty function, which decreases as fast as possible for a strongly consistent estimator, as sample size increases. Hannan-Quin criterion is given by

$$HQ = \ln \hat{\sigma}^2 + 2(p + q) \frac{ln(\ln n)}{n} \quad (6)$$

Hannan and Rissanen (1982) replace the term Ln (1nn) by 1nn to speed up the covergence of HQ. This revised version of HQ, however, was found to overestimate the model orders (Kavaleris 1991).

In 1989, Hurvish and Tsai found that BIC, which was modified from AIC, is not asymptotically efficient. Hence, they suggested a biased corrected version of AIC, known as Akaike's information corrected criterion or AICC. AICC statistic is given by

$$AICC = -2 \ln \textit{Likelihood } (\hat{\phi}, \hat{\theta}, \hat{\sigma}^2) + [2n(p + q + l)]/[n - (p + q) - 2] \quad (7)$$

where $\hat{\phi}$ are estimated autoregressive parameters,
$\hat{\theta}$ are estimated moving average parameters,
$\hat{\sigma}^2$ is estimated variance of white noise,
n is number of observations,
p is order of the autoregressive component,
q is order of the moving average component,

and *Likelihood* $(\hat{\phi}, \hat{\theta}, \hat{\sigma}^2)$ is the likelihood of the data under the Gaussion ARMA model with parameters $(\hat{\phi}, \hat{\theta}, \hat{\sigma}^2)$.

The penalty factors $2n(p + q + 1)/[n - (p + q) - 2]$ and $2(p + q + 1)$, for AICC statistics and AIC statistics respectively, are asymptotically equivalent as $n \to \infty$. Moreover, AICC, as AIC or PE, is asymptotically efficient for autoregressive process. The AICC statistics however, has a more extreme penalty for large order models, which counteract the over fitting nature of the AIC (Brockwell and Davis 1996). Today, the AICC statistics, as its earlier version(AIC), has been widely used as one of the order selection criteria in ARMA time series as well as the lag-length selection criteria in econometric modelling processes. Due to its popularity, Brockwell and Davis (1994) for instance, have included the AICC statistic in their computer software package known as "*Iterative Time Series Modelling (ITSM)*". As the AICC statistics is an important criterion for the selection of order in time series models, its performance must be evaluated. The study hence takes the initiative to explore the probability of minimum AICC criterion in picking up the *true* model based on a simulation study.

SIMULATION STUDY

In this study, a total of 10,000 simulated data series from 10 autoregressive moving average processes were investigated. These processes were AR(1), AR(2), AR(3), AR(4), MA(1), MA(2), ARMA(1,1), ARMA(1,2), ARMA(2,1)and ARMA(2,2). From there, 100 models were formulated in such a way that each process was assigned a number of 10 models. These models are summarized in the Appendix. For illustration, the 10 models for AR(1) process were those with a parameter ϕ value of 0.10. 0.30. 0.50, 0.70, 0.90, -0.30, -0.50, -0.60, -0.80 and -0.95 respectively. Each of these 10 models is in turn replicated into 100 random data series using a different random seed number (less than 10 digits) for each replication. To be consistent in comparison, every random series has 555 observations with a mean value of 111 and unit variance. No element of seasonality or trend is involved in this simulated data. The data series are randomly generated using the "Generation of the Simulated Data" option of the ITSM software.

The process of time series model fitting in this study involves identification of appropriate models, estimation of parameters and validation of the model. In the process of model fitting, ITSM automatically selected a minimum AICC model for each of the data series generated from the AR(1), AR(2), AR(3) and AR(4) processes. As for each of the remaining series, 4 to 9 appropriate models were fitted for model selection purpose. The estimated models are appropriate in the sense that, besides they are stationary and invertible, they are also required to pass the following formal diagnostic tests of randomness.

1. Ljung-Box portmanteau test, which uses the autocorrelations of the residuals to test for the null hyphotesis that the residuals are independently and identically distributed (iid);
2. Mcleod-Li portmanteau test, which tests whether the residuals are from an iid sequence of normally distributed random variables, by using the autocorrelations of the squarred-residuals;
3. Turning point test, which is normality test based on the number of turning points;
4. Different sign test, which is used to detect whether a linear trend (implies non-stationary) is present in the residuals;
5. Rank test, which is also a stationary rest for the residuals.

These test are easily checked by "Tests of Randomness of the Residuals" option in the software mentioned earlier. The order of the Yule-Walker model for the residuals is also estimated by this option, to asses whether the residuals of the each estimated model are compatible with the white plotting the sample autocorrelation function (ACF) and partial autocorrelation function (PACF) are performed by the "Model ACF/PACF" option of ITSM software. The details on these diagnostic tests are available in Brockwell and Davis (1996). Out of a class of appropriate models, the order p and q of the minimum AICC model were recorded for each series.

If the estimated p and q of the minimum AICC model matches the simulated model, we say that the AICC criterion has picked up the correct model. If it failed to pick up the correct model, further investigation was carried out to determine whether over parameterization or under parameterization has occurred. Due to the fact that in the computation of AICC statistics the sum of p and q exceeding sum of the true order p and q, whereas under parameterization happened when sum of the true order p and q exceeding sum of the estimated order p and q. With these definitions, a minimum AICC model might fail to pick up the correct model, due to neither over parameterization not under parameterization, however. For instance, ARMA(1,2), ARMA(3,0) and ARMa(0,3) models were clearly different from ARMA(1,2) model, but neither of them was considered over parameterization or under parameterization. This paradox stemmed from the deficiency in the computation of AICC statistics, which regarded $p + q$ as one term. In this study, these models are treated as misspecified models.

In this study, for every 100 series of the same model, the probability that the minimum AICC model picks up the correct model, denote by P_c was computed as

$$P_c = \frac{\text{number of time "pick up" occurred}}{100}. \tag{8}$$

The probability that the event "over parameterization" happened, *Po was* calculated as

$$P_o = \frac{\text{number of time "over parameterization" occurred}}{100}. \tag{10}$$

Finally, the probability that the event "mis-specification" occurred, P_m was determined by

$$P_m = \frac{\text{number of time "mis-specification" occurred}}{100}. \tag{11}$$

SIMULATION RESULT

Amongst the 10 models of AR(1) process P_c ranged from 0.63 to 0.81 with a mean value of 0.721; P_o ranged from 0.19 to 0.37 with a mean value of 0.268, while P_u ranged from 0 to 0.99 with a mean value of 0.011. This mean that out of all the 1000 series of AR(1) process, the minimum AICC model matches the correct model 721 of the time; over parameterization occurs 268 of the time and under parameterization happens only 11 of the time. The result for AR(1) process and other procesesses in this study was summarized in Table 1. From this criterion, with a probability of picking the true model ranging from 0.366 to 0.795 and a mean value of 0.613. However, changes of over parameterization still exist and in very 100 models, around 17 to 50 models will be over parameterized. As compared to Autoregressive of Moving-Average models, over parameterization was found relatively serious in mixed Autoregressive Moving-Average models, where the AICC statistics could pick up at most 60 percent of the correct models. The AICC statistics in picking up the "mis-specified" model was negligible in only 4 out of 100 models (not shown). This result suggests that

TABLE 1
Summary of simulation's results

No	Process	Correctly estimated			Over parameterization			Under parameterization		
		Low	High	Mean	Low	high	Mean	Low	High	Mean
1	AR(1)	.63	.81	.721	.19	.37	.268	.00	.09	.011
2	AR(2)	.52	.84	.751	.16	.25	.219	.00	.25	.030
3	AR(3)	.60	.79	.714	.19	.32	.255	.00	.16	.031
4	AR(4)	.25	.78	.631	.15	.33	.233	.00	.60	.097
5	MA(1)	.43	.79	.670	.19	.41	.256	.00	.04	.005
6	MA(2)	.56	.84	.733	.16	.44	.265	.00	.00	.000
7	ARMA(1,1)	.20	.87	.601	.11	.80	.358	.00	.13	.013
8	ARMA(1,2)	.45	.74	.594	.26	.55	.406	.00	.00	.000
9	ARMA(2,1)	.01	.84	.320	.11	.71	.302	.00	.84	.246
10	ARMA(2,2)	.01	.65	.393	.22	.82	.413	.00	.62	.116
	Overall	.366	.795	.613	.174	.500	.298	.000	.273	.055

whenever the minimum AICC criterion failed to pick up the *true* model correctly, it was due to over parameterization. This fact that AICC over parameterized could be perceived as supportive to the proponents of parsimonous model such as Box and Jenkins (1976). Hence for any two comparable models, it is always safe to choose the one with lower order p and q.

CONCLUSION

The AICC statistics, as its earlier version (AIC) has been widely used as one of order selection criteria in ARMA time series as well as the lag-length selection criterion in econometric processes. As the AICC statistics is important in ARMA time series modelling and related fields, its performance must be evaluated. This paper evaluates the performance of AICC by determining the probability of the minimum AICC criterion in picking up the true model based on a simulation study. A total of 100 models from 10 ARMA processes were used in this study, with 100 replicants for each model giving to a total of 10,000 data series. The probability if interest was found to be only 0.613, even though we had use considerably large sample size. Hence, the performance of AICC in picking up the true models is expected to decline in the case of smaller sample size, which usually happens in empirical research. In addition, the minimum AICC criterion, which tries to overcome the over parameterization of the minimum AIC criterion, still has the tendency to overestimate the model orders. This implies that applying AICC criterion in either time series modelling or the selection of lag-length for any lag-length sensitive test such as unit root and cointegration test in the related fields would weaken the credibility of the ultimate result.

This study investigation only 10 of the commonly used ARMA (p,q) processes. It could be improved by including more variations of process, especially those with moderately high order, to produce a more influential result. The sample size could also be varied such that the actual performance of the minimum AICC criterion in conjunction with various sample size could be uncovered. A computer search algorithm could also be designed to determine a new empirically sound order selection criterion.

REFERENCES

AKAIKE, H. 1969. Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics* **21**: 243-247.

AKAIKE, H. 1970. Statistical predictor identification. *Annals of the Institute of Statistical Mathematics* **22**: 20-217.

AKAIKE, H. 1973. Information Theory and an Extension of the Maximum Likelihood Principle. In Peron, B.N. and Csaki, F. (eds), *2nd International Symposium in Information Theory*, p. 207-261. Budapest: Akademial Kiodd.

AKAIKE, H. 1979. A Bayesian extension of the minimum AIC procedure of autoregressive model fitting. *Biometrica* **66(2)**: 237-242.

BEVERIDGE, S. and C. OICKLE. 1994. A comparison Box-Jenkins and objective methods for determining the order of a non-seasonal ARMA model. *Journal of Forecasting* **13**: 419-434.

BOX, G. E. P. and G. M. JENKINS. 1976. *Time Series Analysis.* Revised edition. San Francisco: Holden-day.

BROCKWELL, P. J. and R. A. Bavis. 1994. *ITSM for Windows.* New York: Springer.

BROCKWELL, P. J. and R. A. Davis. 1996. *Introduction to Time Series and Forecasting.* New York: Springer.

DE GOOIJER, G., B. ABRAHAM, A. GOULD and L. ROBINSON. 1985. Methods for determining the order of an autoregressive moving average process: a survey. *International Statistical Review* **53**: 301-329.

GEWEKE, J. F. and R. A. MEASE. 1981. Estimating regression of finite but unknown order. *International Economic Review* **22**: 55-77.

HANNAN, E. J. 1980. The estimation of the order of an ARMA process. *Annals of Statistics* **8**: 1071-1081.

HANNAN, E. J. and B. G. QUINN. 1979. The determination of the order of an autoregression. *Journal of Royal Statistical Society* **41(2)**: 190-195.

HANNAN, E. J. and J. RISSANEN. 1982. Recursive estimation of mixed autoregressive moving average order. *Biometrica* **69(1)**: 81-94.

HARVEY, A. C. 1993. *Time Series Model.* 2nd ed. UK: Harvester Wheatsheaf.

HURVISH C., M and C. L. TSAI. 1989. Regression and time series model selection in small samples. *Biometrica* **76**: 297-307.

JONES, R. H. 1975. Fitting autoregressions. *Journal of American Statistics Association* **70**: 590-592.

KAVALIERIS, L. 1991. A note on estimating the dimension of a model. *The Annals of Statistics* **6(2)**: 461-464.

SHIBATA, R. 1976. Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrica* **63(1)**: 117-126.

STOICA, P., P. EYKHOFF, P. JANSEN and T. SÖDERSTRÖM. 1986. Model selection by cross validation. *International Journal of Control* **43**: 1841-1878.

RISSANEN, J. 1978. Modelling by shortest data description. *Automatica* **14**: 467-471.

ISSN: 0128-7680
Pertanika J. Sci. & Technol. **10(1)**: 35-54 (2002)

Cost and Performance Analysis of Integrity Validation Techniques for a Distributed Database

Hamidah Ibrahim
Department of Computer Science
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
43400 UPM, Serdang, Selangor, Malaysia
e-mail: hamidah@fsktm.upm.edu.my

Received: 14 December 2000

ABSTRAK

Satu masalah utama dengan penggunaan kekangan integriti untuk mengawasi integriti pangkalan data yang berubah secara dinamik ialah kos penilaiannya. Kos ini yang berkait dengan prestasi mekanisme pemeriksaan ialah ukuran kuantitatif yang utama yang harus diselia dengan teliti. Kami telah membangunkan satu subsistem kekangan integriti untuk suatu pangkalan data teragih hubungan yang mengandungi beberapa teknik yang diperlukan untuk pemeriksaan kekangan secara efisien, terutamanya di dalam persekitaran teragih di mana pengagihan data adalah bebas kepada domain aplikasi. Di dalam makalah ini, kami akan menunjukkan bagaimana teknik-teknik ini telah dengan efektifnya mengurangkan kos pemeriksaaan kekangan di dalam persekitaran teragih.

ABSTRACT

A principal problem with the use of integrity constraints for monitoring the integrity of a dynamically changing database is their cost of evaluation. This cost which is associated with the performance of the checking mechanisms is the main quantitative measure which has to be supervised carefully. We have developed an integrity constraint subsystem for a relational distributed database (SICSDD) which consists of several techniques that are necessary for efficient constraint checking, particularly in a distributed environment where data distribution is transparent to the application domain. In this paper, we will show how these techniques have effectively reduced the cost of constraint checking in such a distributed environment.

Keywords: Distributed database, integrity constraints, constraint checking

INTRODUCTION

A database state is said to be consistent if the database satisfies a set of constraints, called *semantic integrity constraints.* Integrity constraints specify those configurations of the data that are considered semantically correct. Any update operation (insert, delete or modify) or transaction (sequence of updates) that occurs must not violate these constraints. Thus, a fundamental issue concerning integrity constraints is constraint checking, that is the process of ensuring that the integrity constraints are satisfied after the database has been updated.

Much attention has been paid to the maintenance of integrity in centralized databases over the last decade. A naive approach to constraint checking is to perform the update and then check whether all the integrity constraints are satisfied in the new database state. This method, termed *brute force checking*, is very expensive, impractical and can lead to prohibitive processing costs (Hsu and Imielinskin 1985; Mazumdar 1993; Qian 1989). Enforcement is costly because the evaluation of all integrity constraints requires accessing large amounts of data which are not involved in the database update state transition. Researchers have suggested that constraint checking can be optimized by exploiting the fact that the constraints are known to be satisfied prior to an update. This strategy known as *incremental integrity checking*, avoids redundantly checking constraints that are satisfied in the database before and are not affected by the update operation. It is the basis of most current approaches to integrity checking in databases. Another strategy is to simplify the constraint formulae so that *less data are accessed* in order to determine the truth of a constraint. Under the assumption that the set of initial constraints, *IC*, is known to be satisfied in the state before an update, simplified forms of *IC*, say *IC'*, are constructed such that *IC* is satisfied in the new state if and only if *IC'* is satisfied, and the evaluation cost of *IC'* is less than or equal to the evaluation cost of *IC*. This strategy is referred to as *constraint simplification* and the simplified forms of these constraints are referred to as integrity tests. This approach conforms with the admonition of Nicholas (1982) to concentrate on the problem of finding *good*[1] constraints. Various simplification techniques have been proposed where integrity tests are derived from the syntactic structure of the constraints and the update operations (Gupta and Widom 1993; Hsu and Imielinskin 1985; McCune and Henschen 1989; Nicholas 1982; Simon and Valduriez 1986). Researchers in this area have focussed solely on the derivation of efficient integrity tests, claiming that they are cheaper to enforce and reduce the amount of data accessed, thus reducing the cost of integrity constraint checking. Three different types of integrity test are defined in (McCune and Henschen 1989), namely: *sufficient tests, necessary tests* and *complete tests.*

Although this research effort has yielded fruitful results that have given centralized systems a substantial level of reliability and robustness with respect to the integrity of their data, there has so far been little research carried out on integrity issues for distributed databases. Devising an efficient algorithm for enforcing database integrity against updates is more crucial in a distributed environment. The reasons for this are described in (Barbara and Garcia-Molina 1992; Mazumdar 1993; Qian 1989; Simon and Valduriez 1986). The brute force strategy of checking constraints is worse in the distributed context since the checking would typically require data transfer as well as computation leading to complex algorithms to determine the most efficient approach. Allowing an update to execute with the intention of aborting it at commit time

[1] Good is intended to mean easy to maintain and easy to check (Nicholas 1982)

in the event of constraint violation is also inefficient since rollback and recovery must occur at all sites in which the update participated.

A principal problem with the use of integrity constraints for monitoring the integrity of a dynamically changing database is their cost of evaluation. This cost which is associated with the performance of the checking mechanisms is the main quantitative measure which has to be supervised carefully. Different criteria have been used to assess this performance such as the time to check the validity of the constraints against updates. Generally, an efficient constraint checking strategy tries to minimize the utilization of the computing resources involved during the checking activities. A common goal addressed by previous researchers in this field is to propose constraint simplification strategies which manage to derive a better set of constraints than the initial set. A simplification strategy is said to be efficient if the evaluation of the generated simplified forms has effectively reduced the cost of integrity checking compared to the evaluation of the initial constraints. Based on the following studies (Bernstein and Blaustein 1981; Gupta and Widom 1993; Hsu and Imielinskin 1985; Mazumdar 1993; Nicholas 1982; Qian 1988; Simon and Valduriez 1986), the cost of evaluating an integrity constraint includes the following *main* components: (i) the amount of data accessed (locally or non-locally for a distributed database) – this is related to the checking space of the integrity constraint (Hsu and Imielinskin 1985; Qian 1988); (ii) the amount of data transferred across the network; and (iii) the number of sites involved, which is interrelated with (ii) above. For a centralized database, the main emphasis is on minimizing the amount of data accessed or the checking space. In distributed databases, where many sites are involved, the amount of data transferred across the network and the number of sites involved must be minimized too. Most of the authors cited consider a single cost component due to the difficulty in assigning suitable weights to all cost components.

Bernstein and Blaustein (1981) and Nicholas (1982) used the following two intuitive arguments as a justification that the evaluation of a simplified constraint produced by their algorithm has effectively reduced the cost of integrity checking compared to the evaluation of the corresponding initial constraint: (i) the more constants that are substituted for variables in a given constraint, the more selective is the constraint, and so the easier it should be to evaluate and therefore the cheaper the evaluation; and (ii) the simplified forms are derived on a minimal substate of the database state and so involve less data access. Hsu and Imielinskin (1985) measured the simplicity of an integrity constraint by the notion of its checking space. The *checking space* of a constraint $IC(v_1, v_2, ..., v_n)$ is defined as $v_1 \times v_2 \times ... \times v_n$ where x is the cartesian product operator, and the v_i's are the range variables in the *IC*. A constraint *IC-i* is said to be *simpler* than a constraint *IC-j* if the checking space of *IC-i* is smaller than the checking space of *IC-j*. The checking space of a constraint is a rough measure of the complexity of its evaluation and the number of variables it has to access. This measurement is later used by other authors (Qian 1988). Simon and Valduriez (1986) claimed that their simplification method minimized the cost of integrity checking since only data subject to update are evaluated. Thus they are reducing the checking space by removing data that is known to be unaffected by the change. Mazumdar (1993) proposed a metric called scatter, σ, to capture the amount of non-local access necessary to evaluate a constraint in a distributed database.

We have developed an integrity constraint subsystem for a relational distributed database (SICSDD). The subsystem provides complete functionality and an efficient strategy for constraint enforcement. Complete functionality is attained through a modular and extensible architecture in which several techniques are incorporated. These techniques are necessary to achieve efficient constraint enforcement, particularly in a distributed database. By database distribution we mean that a collection of data which belongs logically to the same system is physically spread over the sites (nodes) of a computer network where intersite data communication is a critical factor affecting the system's performance. In this paper, we will show how the SICSDD techniques have effectively reduced the cost of constraint checking in a distributed environment. We do this by analysing and comparing the generated simplified forms to their respective initial constraints with respect to the amount of data that has to be accessed, the amount of data transferred across the network and the number of sites that may be involved during the evaluation of these constraints/ simplified forms. In general, our strategy reduces the amount of data needing to be accessed since only fragments of relations subject to update are evaluated. The amount of data transferred across the network and the number of sites that may be involved are minimized by evaluating the simplified forms at the target site, i.e. the site where an update is to be performed.

PRELIMINARIES

Our approach has been developed in the context of relational databases (Date 1995), which can be regarded as consisting of two distinct parts, namely: an intensional part and an extensional part. A database is described by a database schema D, which consists of a finite set of relation schemas, $<R_1, R_2, \ldots, R_m>$. A relation schema is denoted by $R(A_1, A_2, \ldots, A_n)$ where R is the name of the relation (predicate) with n-arity and the A_i's are the attributes of R. Let dom(A_i) be the domain values for attribute A_i. Then, an instance of R is a relation $\Re$ which is a finite subset of cartesian product dom(A_1) x ... x dom(A_n). A database instance is a collection of instances for its relation schemas. A relational distributed database schema is described as a quadruple (D, IC, FR, AS) where IC is a finite set of integrity constraints, FR is a finite set of partitioning rules and AS is a finite set of allocation schemas.

The database integrity constraints considered in this paper are *state constraints* which include: domain (*IC-1*), key (*IC-2, IC-3*), referential (*IC-4*) and general semantic integrity constraints (*IC-5, IC-6*). They are expressed in prenex conjunctive normal form with the range restricted property (McCune and Henschen 1989; Nicholas 1982). A conjunct (literal) is an atomic formula of the form $R(u_1, u_2, \ldots, u_k)$ where R is a k-ary relation name and each u_i is either a variable or a constant. A positive atomic formula (positive literal) is denoted by $R(u_1, u_2, \ldots, u_k)$ whilst a negative atomic formula (negative literal) is prefix by $\neg$. An (in)equality is a formula of the form u_1 OP u_2 (prefix with $\neg$ for inequality) where both u_1 and u_2 can be constants or variables and $OP \in \{<, \leq, >, \geq, <>, =\}$.

A set of fragmentation rules, FR, specifies the set of restrictions, C_i, that must be satisfied by each fragment relation $\Re_i$. These rules introduce a new set of integrity constraints and therefore have the same notation as IC. We assume that the fragmentation of relations satisfies the completeness, the disjointness and the reconstructability properties (Ozsu and Valduriez 1991). An allocation schema locates a fragment relation, $\Re_i$, to one or more sites. Throughout this paper the same example *emp_dept* database is used, as given in *Fig. 1*. Here we assume that the relation *emp* is first vertically fragmented into emp_1 on attribute set A_1 = {*eno, ename, eaddress*} and emp_2 on attribute set A_2 = {*eno, dno, ejob, esal*};

Schema: emp(eno,ename,eaddress,dno,ejob,esal)
dept(dno,dname,mgrno,mgrsal)
Integrity Constraints (Global Constraints):
'A specification of valid salary'
IC-1 : $(\forall u \forall v \forall w \forall x \forall y \forall z)(emp(u, v, w, x, y, z) \rightarrow (z > 0))$
'Every employee has a unique eno'
IC-2 : $(\forall u \forall v1 \forall v2 \forall w1 \forall w2 \forall x1 \forall x2 \forall y1 \forall y2 \forall z1 \forall z2)$ $(emp(u, v1, w1, x1, y1, z1) \wedge emp(u, v2, w2, x2, y2, z2)$
$\rightarrow (v1 = v2) \wedge (w1 = w2) \wedge (x1 = x2) \wedge (y1 = y2) \wedge (z1 = z2))$
'Every department has a unique dno'
IC-3 : $(\forall w \forall x1 \forall x2 \forall y1 \forall y2 \forall z1 \forall z2)(dept(w, x1, y1, z1) \wedge dept(w, x2, y2, z2) \rightarrow (x1 = x2) \wedge (y1 = y2) \wedge (z1 = z2))$
'The dno of every tuple in the *emp* relation exists in the *dept* relation'
IC-4 : $(\forall r \forall s \forall t \forall u \forall v \forall w \exists x \exists y \exists z)(emp(r, s, t, u, v, w) \rightarrow dept(u, x, y, z))$
'Every manager in department *D1* earns more than 4000 pounds'
IC-5 : $(\forall w \forall x \forall y \forall z)(dept(w, x, y, z) \wedge (w = D1) \rightarrow (z > 4000))$
'Every employee must earn ≤ to the manager in the same department'
IC-6 : $(\forall r \forall s \forall t \forall u \forall v \forall w \forall x \forall y \forall z)(emp(r, s, t, u, v, w) \wedge dept(u, x, y, z) \rightarrow (w \leq z))$
Fragmentation Rules:
FR-1: $(\forall w \forall x \forall y \forall z)(emp_{21}(w, x, y, z) \rightarrow (x = D1))$
FR-2: $(\forall w \forall x \forall y \forall z)(emp_{22}(w, x, y, z) \rightarrow (x = D2))$
FR-3: $(\forall w \forall x \forall y \forall z)(dept_1(w, x, y, z) \rightarrow (w = D1))$
FR-4: $(\forall w \forall x \forall y \forall z)(dept_2(w, x, y, z) \rightarrow (w = D2))$

Fig. 1: The emp_dept intensional database

emp_2 and *dept* are horizontally fragmented into two fragments emp_{2i} and $dept_i$, respectively, with predicates $dno = D1$ and $dno = D2$.

The key problem in integrity checking is how to efficiently evaluate constraints. An accurate way of measuring the evaluation of the generated fragment constraints/simplified forms is a high priority so that they can be compared fairly and realistically. In this paper, we will evaluate the derived set of fragment constraints/simplified forms with respect to the following components. We use the symbol $C(R_1, R_2, \ldots, R_n)$ to denote the set of relations or fragment relations specified in the constraint/simplified form C.

A – provides an estimate of the amount of data accessed, which is related to the number and the size of the relations or fragment relations specified in a given constraint/simplified form. This measurement indirectly indicates the size of the checking space. It is based on the following formula: $A_{C(\Re 1, \Re 1, \ldots, \Re n)} = \delta\Re 1 + \delta\Re 2 + \ldots + \delta\Re n$ where the $\delta\Re i$'s are the relations or fragment relations specified in the C and $\delta\Re i$ is the size of $\Re i$. For a compound constraint $C = \Lambda^{m}_{i=1} C_i$, where C_i is a simple constraint, the amount of data accessed is measured as follows: $\Sigma^{m}_{i=1} A_{ci}$. For a disjunction of constraints, $C = V^{m}_{i=1} C_i$, the amount of data accessed is measured as follows: $[A_{min}, \Sigma^{m}_{i=1} A_{ci}]$ which is a range where A_{min} ($\Sigma^{m}_{i=1} A_{ci}$, respectively) is the minimum (maximum, respectively) amount of data that might need to be accessed.

τ – provides an estimate of the amount of data transferred across the network. It is measured based on the following formula, $\tau = \Sigma^{n}_{i=1} dt_i$, where dt_i is the amount of data transferred from site i to the target site.

σ - gives a rough measurement of the amount of non-local access necessary to evaluate a constraint/simplified form, and is taken from Mazumdar (1993). This is measured by analysing the number of sites that might be involved in validating the constraint. In general, the formula applied is as follows: $\sigma_C = |site(C)|$ where σ_C is the locality of C, i.e. the number of sites involved, and $site(C)$ is obtained by the following formula.

FOR each distinct relation or fragment relation, fr_i for $i = \{1, 2, \ldots, k\}$ in C DO
site$(C) = \{j : j \in$ sites Λ fr_i is allocated at site j where sites $\in \{1, 2, \ldots, n\}\}$

For a compound constraint $C = \Lambda^{m}_{i=1} C_i$, where C_i is a simple constraint, the locality of these constraints is simply obtained from the maximum number of sites that are involved in evaluating one of the constraints, i.e. $\sigma_C = max(\sigma_{C1}, \sigma_{C2}, \ldots, \sigma_{Cm})$. The locality of an integrity test, given an update operation, gives a rough measurement of the amount of non-local access necessary for verifying if the integrity test is being satisfied or violated by the update operation. This is measured by analysing the number of sites that might be involved in validating the test. In general, the formula applied is as follows:

$$\sigma_{Ti} = \begin{cases} |site(T_i)| & if\, site(U_i) \subseteq site(T_i) \quad (1) \\ |site(T_i)| + 1 & otherwise \end{cases} \tag{1}$$

where σ_{Ti} is the locality of an integrity test, T_i, with respect to a given update operation, U_i, i.e. the number of sites involved. The site(T_i) and site(U_i) are obtained by the following formula.

FOR each distinct relation or fragment relation, fr_i for $i = \{1, 2, \ldots, k\}$ in T_i DO

$site(T_i) = \{j : j \in$ sites $\wedge$ fr_i is allocated at site j where sites $\in \{1, 2, ..., n\}\}$
$site(U_i) = \{l : l \in$ sites[2] $\wedge$ fr_{ui} is allocated at site l where sites $\in \{1, 2, ..., n\}$ and fr_{ui} is a fragment relation specified in $U_i\}$

Table 1 shows information relating to the *emp_dept* database. This information includes each relation $\Re$, its size $\delta\Re$, its fragment relation $\Re_i$ and their sizes $\delta\Re_i$. The fragment relations are derived based on the fragmentation schemas given earlier in this section. Case I and Case II in the table represent the sites where

TABLE 1
The *emp_dept's* relations and fragment relations

$\Re$	$\delta\Re$	$\Re_i$	$\delta\Re_i$	Case I[a]	Case II[a]
emp	E[b]	emp_1	e_1	S_0	S_0
		emp_{21}	e_{21}	S_1	S_1
		emp_{22}	e_{22}	S_2	S_2
dept	D[c]	$dept_1$	d_1	S_1	S_3
		$dept_2$	d_2	S_2	S_4

[a] $S_i \neq S_j$ where $i \neq j$
[b] $E = e_1 + \Sigma^2_{i=1} e_{2i} - \delta_{key}$ where δ_{key} is the size of the repeated primary key values
[c] $D = \Sigma^2_{i=1} d_i$

the fragment relations are allocated. Here Case I represents the reasonable case where the fragment relations constructed based on the same fragmentation rules are allocated to the same site. It is called the reasonable case as it should minimize the network traffic. Case II represents a worst case where each fragment relation is allocated to a different site of the network. It is a worst case in that network traffic is involved in any constraint evaluation that involves more than one fragment.

OUR OPTIMIZATION TECHNIQUES AND THEIR PERFORMANCES

The high execution cost of constraint enforcement is one of the major problems in the field of constraint handling (Grefen 1993). This cost can be substantially reduced not only by applying an efficient enforcement strategy but also by generating/evaluating a good set of integrity constraints. The techniques that are incorporated into our system seek to derive *efficient* sets of fragment constraints and a range of possible local tests [3]. Our techniques are identified

[2] As we assume that there is no replication of fragments across the network, for a given update operation, U_i there will be only one target site, i.e. the site where the update is to be performed.
[3] An integrity test is a local test if it can be evaluated at a single site, i.e. at the site where the update is to be performed, and a global test otherwise.

as *constraint preprocessing, constraint distribution* and *integrity test generation.* In this section, we will evaluate the derived set of fragment constraints/simplified forms with respect to the components described in the previous section.

Constraint Preprocessing Techniques

There are five steps that are applied which are performed by the following procedures (we use the notation X^{-} to indicate a finite set of variables, $x_1, x_2, \ldots, x_k$; and x_{key}, x_{join} and x_{ref} are the key, join and reference attribute, respectively). In general, the amount of data accessed for evaluating the set of fragment constraints, A_{FC-i}, derived by the procedures embodied in the constraint preprocessing techniques is less than the amount of data accessed for evaluating the initial constraint, A_{IC-i}, i.e. $A_{FC-i} < A_{IC-i}$. $\delta\Re_{min}$ is *minimum* $(\delta\Re_1, \delta\Re_2, \ldots, \delta\Re_n)$.

constraint_transformation_procedure

This procedure transforms the constraints specification at the relational level (global constraints) into a constraints specification at the fragment level (fragment constraints). The objective of the transformation is to obtain a specification of constraints that can be straightforwardly used for constructing efficient enforcement algorithms. At this stage, the transformations are restricted to logically equivalent transformations, without considering any reformulation of the original constraints. Initially, each occurrence of a relation $\Re$ in a constraint is replaced by its n fragment relation, $\Re_i$. The transformation is one to many, which means that given a global constraint, the result of the transformation is a logically equivalent set of fragment constraints. There are four transformation rules which are applied during this process. These rules analyse the patterns in the prefixes of a constraint and the types of fragmentation that are being applied to the global relations. The proofs of these rules can be found in (Qian 1989) and are therefore omitted here.

Horizontal transformation rules:

i. $(\forall X^{-})(R(X^{-}) \rightarrow P(X^{-})) \equiv \Lambda^{n}_{i=1}(\forall X^{-})(R_i(X^{-})\ P(X^{-}))$

ii. $(\exists X^{-})(R(X^{-}) \wedge P(X^{-})) \equiv \mathrm{V}^{n}_{i=1}(\exists X^{-})(R_i(X^{-}) \wedge P(X^{-}))$

Vertical transformation rules:

i. $(\forall X^{-})(R(X^{-}) \rightarrow P(X^{-})) \equiv (\forall x_{key} \forall X^{-}_1 \ldots \forall X^{-}_n)(R_1(x_{key}, X^{-}_1) \wedge \ldots \wedge R_n(x_{key}, X^{-}_n) \rightarrow P(x_{key}, X^{-}_1, \ldots, X^{-}_n))$

ii. $(\exists X^{-})(R(X^{-}) \wedge P(X^{-})) \equiv (\exists x_{key} \exists X^{-}_1 \ldots \exists X^{-}_n)(R_1(x_{key}, X^{-}_1) \wedge \ldots \wedge R_n(x_{key}, X^{-}_n) \wedge P(x_{key}, X^{-}_1, \ldots, X^{-}_n))$

To obtain a logically equivalent set of fragment constraints from a given global constraint which contains a relation $\Re$ and is fragmented by a mixed fragmentation, we repeatedly apply horizontal and vertical transformation rules in the same sequence of horizontal and vertical fragmentations to the global constraint to produce the fragment constraints. Example: The integrity constraint *IC-4* is transformed into the following set of fragment constraints.

$$FC\text{-}4_t : \bigwedge_{i=1}^{2} \bigvee_{j=1}^{2} (\forall t \forall u \forall v \forall w \exists x \exists y \exists z)(emp_{2i}(t, u, v, w) \rightarrow dept_j(u, x, y, z))$$

As expected, evaluating the sets of fragment constraints derived by the constraint_transformation_procedure without further optimization is inefficient (with respect to the amount of data accessed) since the evaluation of these constraints generally will access all the fragment relations specified in these constraints. This is similar to the brute force strategy, but is only the initial stage of our application.

simplification_procedure

This procedure, which uses knowledge about the fragmentation of relations, reduces the number of fragment relations involved in the evaluation of a constraint. A set of fragment constraints derived from a global constraint can be simplified if the relations specified in the global constraint are fragmented on a join/reference attribute using the same fragmentation algorithm. Instead of deriving all combinations of fragment constraints, only compatible fragment constraints are constructed. Given global relations $\mathfrak{R}$ and S which are fragmented on a join/reference attribute into n and m fragment relations, respectively on the same fragmentation rules, then the following *simplification rules* are applied.

i. $\bigwedge_{i=1}^{n} \bigvee_{j=1}^{m} (\forall x_{ref} \forall X^{-} \exists Y^{-})(R_i(x_{ref}, X^{-}) \rightarrow S_j(x_{ref}, Y^{-})) \equiv \bigwedge_{i=1}^{n} (\forall x_{ref} \forall X^{-} \exists Y^{-}) (R_i(x_{ref}, X^{-}) \rightarrow S_i(x_{ref}, Y^{-}))$

ii. $\bigwedge_{i=1}^{n} \bigwedge_{j=1}^{m} (\forall x_{join} \forall X^{-} \forall Y^{-})(R_i(x_{join}, X^{-}) \wedge S_j(x_{join}, Y^{-}) \ldots \rightarrow \ldots) \equiv \bigwedge_{i=1}^{n} (\forall x_{join} \forall X^{-} \forall Y^{-})(R_i(x_{join}, X^{-}) \wedge S_i(x_{join}, Y^{-}) \ldots \rightarrow \ldots)$

iii. $\bigvee_{i=1}^{n} \bigvee_{j=1}^{m} (\exists x_{join} \exists X^{-} \exists Y^{-})(R_i(x_{join}, X^{-}) \wedge S_j(x_{join}, Y^{-}) \ldots) \equiv \bigvee_{i=1}^{n} (\exists x_{join} \exists X^{-} \exists Y^{-})(R_i(x_{join}, X^{-}) \wedge S_i(x_{join}, Y^{-}) \ldots)$

Example: $FC\text{-}4_t$ above can be simplified since both relations *emp* and *dept* are fragmented on the reference attribute, i.e. *dno*, using the same fragmentation rules. The simplified set of fragment constraints are as shown below.

$$FC\text{-}4_{si} : \bigwedge_{i=1}^{2} (\forall t \forall u \forall v \forall w \exists x \exists y \exists z)(emp_{2i}(t, u, v, w) \rightarrow dept_i(u, x, y, z))$$

The simplification_procedure simplifies a set of fragment constraints which reduces the number of joins/references between fragment relations required in the definition of the set of fragment constraints.

subsumption_procedure

This procedure attempts to obtain an improved set of fragment constraints by removing redundant fragment constraints from a set. A redundant fragment constraint is a constraint that can be implied (syntactically) from other existing fragment constraints. Thus, deleting a redundant fragment constraint from its set does not affect the unsatisfiability/satisfiability of the set since the truth of this constraint can be inferred from the truth of its identical or subsumed pair.

Even though a redundant set of constraints is semantically correct, excluding redundancy can improve the enforcement time. Example: Consider the following set of fragment constraints derived from *IC-3* by applying the transformation process constraint_transformation_procedure.

FC-3t: $\wedge_{i=1}^{2} \wedge_{j=1}^{2}$ $(\forall w \forall x1 \forall x2 \forall y1 \forall y2 \forall z1 \forall z2)$ $(dept_i(w, x1, y1, z1) \wedge dept_j(w, x2, y2, z2) \rightarrow (x1 = x2) \wedge (y1 = y2) \wedge (z1 = z2))$

The sets $[(\forall w \forall x1 \forall x2 \forall y1 \forall y2 \forall z1 \forall z2)(dept_i(w, x1, y1, z1) \wedge dept_j(w, x2, y2, z2) \rightarrow (x1 = x2) \wedge (y1 = y2) \wedge (z1 = z2))]$ and $[(\forall w \forall x1 \forall x2 \forall y1 \forall y2 \forall z1 \forall z2)(dept_j(w, x1, y1, z1) \wedge dept_i(w, x2, y2, z2) \rightarrow (x1 = x2) \wedge (y1 = y2) \wedge (z1 = z2))]$ for $i \in \{1, 2\}$ and $j \in \{1, 2\}$ are mutually redundant. Removing the redundant fragment constraints will generate the following set of fragment constraints:

FC-3$_{su}$: $\wedge_{i=1}^{2} V_{j=i}^{2}$ $(\forall w \forall x1 \forall x2 \forall y1 \forall y2 \forall z1 \forall z2)$ $(dept_i(w, x1, y1, z1) \wedge dept_j(w, x2, y2, z2) \rightarrow (x1 = x2) \wedge (y1 = y2) \wedge (z1 = z2))$

The subsumption_procedure removes any redundant fragment constraints from a given set of fragment constraints, i.e. $\wedge_{j=1}^{n} fc_j$ or $V_{j=1}^{m} fc_j$ or a set of fragment constraints constructed by both. Obviously, this set of fragment constraints is derived from the horizontal/mixed transformation rules.

contradiction_procedure

This procedure removes the fragment constraints produced by the constraint_transformation_procedure which contradict the fragmentation rules, i.e. it removes fragment constraints that are never satisfied. The contradiction_procedure designed by us is a specific-purpose theorem prover that resembles a resolution-based theorem prover similar to Henschen *et al.* (1984) and McCarrol (1995). It is mainly designed for investigating if a fragment constraint violates one of the existing fragmentation rules. Example: The fragment constraint *FC-3$_{su}$* above, when $i \neq j$, contradicts the fragmentation rules *FR-3* and *FR-4*.

The contradiction_procedure eliminates fragment constraints from their sets if they contradict one of the existing fragmentation rules, i.e. $\wedge^{n}_{j=1} fc_j$ or $V^{m}_{j=1} fc_j$ or a set of fragment constraints constructed by both. Obviously, this set of fragment constraints is derived from the horizontal/mixed transformation rules.

reformulation_procedure

The above procedures (the simplification_procedure, the subsumption_procedure and the contradiction_procedure) use knowledge about data fragmentation and analyse the syntax of the constraints to either remove inessential constraints (a redundant constraint or a constraint which contradicts the fragmentation rules) from the constraints set or reduce the scope of the fragment relations specified in a constraint (the simplification-procedure). These procedures, which are based on syntactic criteria, do not check the possible occurrence of redundant semantics in the constraint set. The reformulation_procedure, which is based on semantic criteria, attempts to (i) remove redundant semantic constructs that may exist, and (ii) reformulate the set of fragment constraints derived so far into alternative forms which can be either an *antecedent* [4] or an *equivalent* form. In our approach, removing the redundant semantic constructs from a given fragment constraint is achieved by applying the substitution and absorption rules. This strategy makes the constraint easier to evaluate because more constants are substituted for the variables in the constraint which makes the constraint more selective. Example: Consider the following fragment constraint derived from *IC-5* by applying the transformation (constraint_transformation_procedure) and the contradiction (contradiction_procedure) processes.

$$FC\text{-}5_l : (\forall w \forall x \forall y \forall z)(dept_l(w, x, y, z) \wedge (w = \mathrm{D1}) \rightarrow (z > 4000))$$

The literal $w = D1$ is a redundant construct as this can be implied from *FR-3*. Removing this construct will generate the following fragment constraint.

$$FC\text{-}5_r : (\forall w \forall x \forall y \forall z)(dept_l(w, x, y, z) \rightarrow (z > 4000))$$

The fragment constraint $FC\text{-}6_{tl}$ below which is derived from *IC-6* can be reformulated as *FC-6'* by using $FC\text{-}5_r$ derived above.

$$FC\text{-}6_{tl} : (\forall t \forall u \forall v \forall w \forall x \forall y \forall z)(emp_{2l}(t, u, v, w) \wedge dept_l(u, x, y, z) \rightarrow (w \leq z))$$
$$FC\text{-}6' : (\forall t \forall u \forall v \forall w)(emp_{2l}(t, u, v, w) \rightarrow (w \leq 4000))$$

Since we require that a fragment constraint which is derived by the reformulation process must be cheaper compared to the initial constraint, therefore the amount of data accessed for evaluating the derived constraint must be less than or equal to the amount of data accessed for evaluating the initial constraint.

Fig. 2 lists the sets of fragment constraints derived after applying the constraint preprocessing techniques to the initial constraints. Each *FC-i* is a semantically equivalent set to its initial constraint *IC-i* (except for *FC-6'* which is an antecedent of $FC\text{-}6_{tl}$).

Table 2 illustrates the amount of data needing to be accessed, A, by the sets of fragment constraints given in *Fig. 2* (except for *FC-6'*) which are derived by the procedures embodied in the constraint preprocessing techniques. In most cases the derived set of fragment constraints is better than the initial constraint which is similar to the brute force strategy, i.e. $A_{FC\text{-}i} < A_{IC\text{-}i}$, as shown by *FC-1*, *FC-*

[4] If an antecedent of a fragment constraint is satisfied this implies that the fragment constraint is satisfied but if the antecedent is falsified, then the fragment constraint has to be evaluated.

3, FC-4, FC-5 and *FC-6* in Table 2. $A_{FC\text{-}i} > A_{IC\text{-}i}$ only when $A_{IC\text{-}i}$ is specified over the attribute(s) of a global relation $\Re$ which is associated with all of its fragment relations which are constructed by a vertical fragmentation, as shown by *FC-2* in Table 2.

$FC\text{-}1 : \wedge^2_{i=1} (\forall w \forall x \forall y \forall z)\ (emp_{2i}(w, x, y, z) \rightarrow (z > 0))$

$FC\text{-}2 : \wedge^2_{i=1} (\forall u \forall v1 \forall v2 \forall w1 \forall w2 \forall x \forall y1 \forall y2 \forall z1 \forall z2)\ (emp_1(u, v1, w1) \wedge emp_{2i}(u, x, y1, z1) \wedge (emp_1(u, v2, w2) \wedge emp_{2i}(u, x, y2, z2) \rightarrow (v1 = v2) \wedge (w1 = w2) \wedge (y1 = y2) \wedge (z1 = z2))$ and $(\forall u \forall x1 \forall x2 \forall y1 \forall y2 \forall z1 \forall z2)\ (emp_{21}(u, x1, y1, z1) \rightarrow \neg(emp_{22}(u, x2, y2, z2))$

$FC\text{-}3 : \wedge^2_{i=1} (\forall w \forall x1 \forall x2 \forall y1 \forall y2 \forall z1 \forall z2)(dept_i(w, x1, y1, z1) \wedge dept_i(w, x2, y2, z2) \rightarrow (x1 = x2) \wedge (y1 = y2) \wedge (z1 = z2))$

$FC\text{-}4 : \wedge^2_{i=1} (\forall t \forall u \forall v \forall w \exists x \exists y \exists z)(emp_{2i}(t, u, v, w) \rightarrow dept_i(u, x, y, z))$

$FC\text{-}5 : (\forall w \forall x \forall y \forall z)(dept_1(w, x, y, z) \rightarrow (z > 4000))$

$FC\text{-}6 : \wedge^2_{i=1} (\forall t \forall u \forall v \forall w \forall x \forall y \forall z)(emp_{2i}(t, u, v, w) \wedge dept_i(u, x, y, z) \rightarrow (w \leq z))$

$FC\text{-}6' : (\forall t \forall u \forall v \forall w)\ (emp_{21}(t, u, v, w) \rightarrow (w \leq 4000))$

Fig. 2: The sets of fragment constraints derived by the constraint preprocessing techniques with respect to the global constraints and fragmentation rules given in Fig. 1

TABLE 2
Estimation of the amount of data accessed – the constraint preprocessing techniques

IC-i	$A_{IC\text{-}i}$	*FC-i*	$A_{FC\text{-}i}$	Comment
IC-1	*E*	*FC-1*	$\Sigma^2_{i=1} e_{2i}$	$A_{FC\text{-}1} < A_{IC\text{-}1}$
IC-2	*E + E*	*FC-2*	$\Sigma^2_{i=1}(2e_1 + 2e_{2i}) + e_{21} + e_{22}$	$A_{FC\text{-}2} > A_{IC\text{-}2}$
IC-3	*D + D*	*FC-3*	$\Sigma^2_{i=1} d_i + d_i$	$A_{FC\text{-}3} < A_{IC\text{-}3}$
IC-4	*E + D*	*FC-4*	$\Sigma^2_{i=1} e_{2i} + d_i$	$A_{FC\text{-}4} < A_{IC\text{-}4}$
IC-5	*D*	*FC-5*	d_1	$A_{FC\text{-}5} < A_{IC\text{-}5}$
IC-6	*E + D*	*FC-6*	$\Sigma^2_{i=1} e_{2i} + d_i$	$A_{FC\text{-}6} < A_{IC\text{-}6}$

The amount of data accessed by *FC-i* is determined by the assumption that all fragment constraints in *FC-i* are evaluated. But given an update operation affecting a fragment relation $\Re_j$, only the fragment constraints in *FC-i* containing $\Re_j$ in their specification should be evaluated. Since we assume that fragmentations satisfy the disjointness property, only a subset of *FC-i* is evaluated, i.e. the amount of data accessed is less than those presented in this section. Also, we assume that the worst case, i.e. $A_{FC\text{-}i} > A_{IC\text{-}i}$, is a rare case since in reality the fragmentation strategies are chosen in such a way that their effect on the integrity constraints will result in efficient constraint checking.

Constraint Distribution Techniques

The fragment constraints constructed so far involve data stored at different network sites. Because the complexity of enforcing constraints is directly related to both the number of constraints in the constraint set and the number of sites involved, our objective in this phase is to reduce the number of constraints allocated to each site for execution at that site. Distributing the whole set of fragment constraints to every site is not cost effective since not all fragment constraints are affected by an update and so sites may not be affected by particular updates.

These techniques reduce the number of constraints allocated to each site by allocating a fragment constraint to a site if and only if there is a fragment relation at that site which is mentioned in the constraint, so that whenever an update occurs at a site, the validation of the fragment constraints at that site implies the global validity of the update. Consequently, these techniques intend to allocate each fragment constraint to a site or minimal number of sites and relieve the irrelevant sites from the computation of certain sets of fragment constraints.

The difference between our approach and Qian's approach (1989) is that the distribution techniques in our approach are applied to the end result of the constraint preprocessing techniques while the distribution techniques in Qian's approach are applied directly to the set of fragment constraints derived by her transformation rules. Our approach is more efficient than hers, since in her approach redundant processing may result during the optimization of the distributed fragment constraints. As a simple example, consider a situation where a fragment constraint requires a join between two fragment relations which are allocated to different sites of the network and this join is deduced to be empty. In our approach, this situation which is detected by the constraint preprocessing techniques will cause the fragment constraint to be eliminated from the set (i.e. no distribution is required). However, in Qian's approach this fragment constraint is distributed to both sites and the optimization of the distributed fragment constraints which is carried out at both sites will result in eliminating both distributed fragment constraints from those sites.

Table 3 shows the effect of the constraint distribution techniques on the derived sets of fragment constraints with respect to σ, which gives a rough measurement of the amount of non-local access required in constraint enforcement. As shown in the table, most of the derived sets of fragment constraints can be evaluated at a single site. Even in the worst case where each fragment is allocated to different sites of the network, the number of sites involved in evaluating the derived sets of fragment constraints is less (in most cases, generally) than the number of sites involved in evaluating the initial constraints (shown by columns Case II of Table 3).

TABLE 3
The reduction in the scatter metric

IC-i	Case I	Case II	*FC-i*	Case I	Case II
IC-1	3	3	*FC-1*	1	1
IC-2	3	3	*FC-2*	3	3
IC-3	2	2	*FC-3*	1	1
IC-4	3	5	*FC-4*	1	2
IC-5	2	2	*FC-5*	1	1
IC-6	3	5	*FC-6*	1	2

Integrity Test Generation Techniques

These techniques generate integrity tests from the syntactic structure of the constraints and the update operations. The algorithms applied for deriving these tests use the substitution, subsumption and absorption rules, and are closely related to Nicholas (1982). These algorithms can be found in Ibrahim *et al.* (1996) and are therefore omitted here.

In a distributed database, four types of integrity test can be identified. They are global post-tests, local post-tests, global pre-tests and local pre-tests (Ibrahim *et al.* 1998). We view local pre-tests as more effective in a distributed database since: (i) only a single site is involved in evaluating the local tests, i.e. the site where the update is to be performed; (ii) as they are evaluated at a target site, this avoids remote reading and the amount of data transferred across the network is minimized – in fact, no data transfer across the network is required (Gupta and Widom 1993); and (iii) they are evaluated before the update is performed – this avoids the need to undo (rollback and recover from) an update in the event of constraint violation, and this reduces the overhead cost of checking integrity (Simon and Valduriez 1986).

The integrity test generation techniques, which derive integrity tests (simplified forms) for the set of fragment constraints, further reduce the amount of data that has to be accessed, A, during the evaluation of these tests. This is illustrated by Table 5. (Table 4 presents the tests constructed for a given update operation and the set of fragment constraints, *FC-i*, given in *Fig. 2*). For example, the domain constraint *IC-1*, which initially required E amount of data to be accessed (see column $A_{IC\text{-}i}$ of Table 2), is reduced to 0 (see column A_{T_i} of Table 5), i.e. no data access is required at all; the referential integrity constraint *IC-4*, which initially required $E + D$ amount of data to be accessed (see column

TABLE 4

The test constructed for a given update operation and the set of fragment constraints, *FC-i*, given in *Fig. 2*

FC-i	INSERT	*Test*, T_i
FC-1	$emp_{2i}(a, b, c, d)$	1. $d > 0$
FC-2	$emp_{2i}(a, b, c, d)$	2. $(\forall v1 \forall v2 \forall w1 \forall w2 \forall x1 \forall y1 \forall z1)(\neg emp_1(a, v1, w1)$ $\vee \neg emp_1(a, v2, w2) \vee \neg emp_{2i}(a, x1, y1, z1) \vee [(v1 = v2)$ $\wedge (w1 = w2) \wedge (x1 = b) \wedge (y1 = c) \wedge (z1 = d)])$ and $(\forall x2 \forall y2 \forall z2)(\neg emp_{2j}(a, x2, y2, z2))$
	$emp_{2i}(a, b, c, d)$	3. $(\forall x1 \forall y1 \forall z1)(\neg emp_{2i}(a, x1, y1, z1))$
FC-3	$dept_i(a, b, c, d)$	4. $(\forall x1 \forall y1 \forall z1)(\neg dept_i(a, x1, y1, z1) \vee [(x1 = b) \wedge (y1 = c)$ $\wedge (z1 = d)])$
	$dept_i(a, b, c, d)$	5. $(\forall x1 \forall y1 \forall z1)(\neg dept_i(a, x1, y1, z1))$
FC-4	$emp_{2i}(a, b, c, d)$	6. $(\exists x \exists y \exists z)(dept_i(b, x, y, z))$
	$emp_{2i}(a, b, c, d)$	7. $(\exists t \exists v \exists w)(emp_{2i}(t, b, v, w))$
FC-5	$dept_l(a, b, c, d)$	8. $d > 4000$
FC-6	$emp_{2i}(a, b, c, d)$	9. $(\forall x \forall y \forall z)(\neg dept_i(b, x, y, z) \vee (d \leq z))$
	$emp_{2i}(a, b, c, d)$	10. $(\exists t \exists v \exists w)(emp_{2i}(t, b, v, w) \wedge (w \leq d))$

A_{IC_i} of Table 2), is reduced to $\sum_{i=1}^{2} e_{2i} + d_i$ (see column A_{FC_i} of Table 2) by the constraint preprocessing techniques and further reduced to d_i (see column A_{Ti} for test 6 of Table 5) or e_{2i} (see column A_{Ti} for test 7 of Table 5) by the integrity test generation techniques.

With respect to the scatter metric, σ, the number of sites involved in evaluating an integrity test for a local fragment constraint is 1 and the number of sites involved in evaluating an integrity test for a non-local fragment constraint is normally more than 1. A local fragment constraint is one in which all fragment relations specified in the constraint are allocated to the same site. For the case of non-local fragment constraints, the number of sites involved can be reduced to 1 by applying the integrity test generation techniques as shown in Table 5. For example, test 9 of *FC-6* for Case II (see column σ_{II} of Table 5) involves two sites, while test 10 of *FC-6* (see column σ_{II} of Table 5) for the same case involves a single site. As most of the work is being carried out at a single site, therefore the amount of data transferred across the network is minimized i.e. $\tau \approx 0$.

The integrity tests listed in Table 4 are generated from the sets of fragment constraints constructed by the fragmentation rules given in *Fig. 1*. As constraint checking is strongly influenced by the fragmentation rules used to construct the fragment relations, it is important to see and analyse the effect of applying different fragmentation rules on the derived integrity tests. This is discussed below. Assume that the fragment emp_{2i} is horizontally fragmented into two fragments emp_{2i} with some predicates different from those given in *Fig. 1*, i.e. the global relations are fragmented based on different fragmentation rules. The

TABLE 5
The reduction in the amount of data accessed, the scatter metric and the amount of data transferred across the network for the integrity tests given in Table 4

Test, T_i	A_{Ti}	σ_I^a	σ_{II}^b	τ_I^c	τ_{II}^d
1.	0		1	0	0
2.	$e_1 + e_1 + e_{2i} + e_{2j}$		3	$e_1 + e_{2j}$	$e_1 + e_{2j}$
3.	e_{2i}		1	0	0
4.	d_i		1	0	0
5.	d_i		1	0	0
6.	d_i		2	0	d_i
7.	e_{2i}		1	0	0
8.	0		1	0	0
9.	d_i		2	0	d_i
10.	e_{2i}		1	0	0

[a] σ of test T_i for Case I.
[b] σ of test T_i for Case II.
[c] τ of test T_i for Case I.
[d] τ of test T_i for Case II.

sets of fragment constraints derived after applying the constraint preprocessing techniques to their respective initial constraints are as shown in *Fig. 3*. Only the derived forms for *IC-4* (represented by $FC\text{-}4_s$) and *IC-6* (represented by $FC\text{-}6_s$) are shown, as the rest of the derived sets of fragment constraints remain the same.

$FC\text{-}4_s: \Lambda^2_{i=1} V^2_{j=1} (\forall t \forall u \forall v \forall w \exists x \exists y \exists z)\ (emp_{2i}(t, u, v, w) \rightarrow dept_j(u, x, y, z))$
$FC\text{-}6_s: \Lambda^2_{i=1} \Lambda^2_{j=1} (\forall t \forall u \forall v \forall w \forall x \forall y \forall z)\ (emp_{2i}(t, u, v, w) \wedge dept_j(u, x, y, z) \rightarrow (w \leq z))$

Fig.3: The sets of fragment constraints derived by the constraint preprocessing techniques when different fragmentation rules are applied

Table 7 shows the effectiveness of the integrity test generation techniques when applied to the sets of fragment constraints derived in *Fig. 3* with respect to A, σ and τ. (Table 6 presents the tests constructed for a given update operation and the set of fragment constraints, *FC-i*, given in *Fig. 3*). From this simple example, it is obvious that constraint checking is strongly influenced by the type of fragmentation and allocation used. Although the resulting simplified forms shown in Table 7 are not as efficient as those presented in Table 5, they are still better than the initial constraints with respect to A, σ and τ. The integrity test generation techniques formulate local simplified forms (shown by tests 12 and 14 in Table 7) which are cheaper than their alternative forms (tests 11 and 13, respectively in Table 7).

TABLE 6
The test constructed for a given update operation and the set of fragment constraints, *FC-i*, given in *Fig. 3*

FC-i	INSERT	Test, T_i
$FC\text{-}4_s$	emp_{2i}	11. $V^2_{j=1}(\exists x \exists y \exists z)\ (dept_j(b, x, y, z))$
	(a, b, c, d)	12. $(\exists t \exists v \exists w)(emp_{2i}(t, b, v, w))$
$FC\text{-}6_s$	emp_{2i}	13. $\Lambda^2_{j=1}(\forall x \forall y \forall z)(dept_j(b, x, y, z) \vee (d \leq z))$
	(a, b, c, d)	14. $(\exists t \exists v \exists w)(emp_{2i}(t, b, v, w) \wedge (w \leq d))$

TABLE 7
The reduction in the amount of data accessed, the scatter metric and the amount of data transferred across the network for the integrity tests given in Table 6

Test, T_i	A_{Ti}	σ_I	σ_{II}	τ_I	τ_{II}
11.	$[d_{min}, \Sigma^2_{j=1} d_j]$	2	3	$[d_{min}, \Sigma^2_{j=1 \text{ and } j<>i} d_j]$	$[d_{min}, \sum_{j=1}^{2} d_j]$
12.	e_{2i}	1	1	0	0
13.	$\Sigma^2_{j=1} d_j$	2	3	$\Sigma^2_{j=1 \text{ and } j<>i} d_j$	$\Sigma^2_{j=1} d_j$
14.	e_{2i}	1	1	0	0

The sufficient tests, which are normally local tests (i.e. $\sigma = 1$), are cheaper in a distributed environment (Gupta and Widom 1993; Mazumdar 1993; Qian 1989; Simon and Valduriez 1986) where the cost of accessing remote data for verifying the consistency of a database is a critical factor influencing the performance of the system (Simon and Valduriez 1986).

Tables 5 and 7 illustrate the improvement in performance gained when using our algorithm rather than Nicolas's algorithm (1982) and other techniques proposed for centralized database. In these tables, the tests 1, 2, 4, 6, 8, 9, 11 and 13 are generated by applying Nicolas's algorithm, while tests 1, 3, 5, 7, 8, 10, 12 and 14 are generated by our algorithm. Comparing those results, most of the tests generated by our algorithm are better than their alternative tests generated by Nicolas's algorithm with respect to the amount of data transferred across the network, τ and the number of sites involved, σ, i.e. the tests generated by our approach can be evaluated at a single site. This is shown in tables 5 and 7, where the tests 3, 7, 10, 12 and 14 are better than their alternative tests 2, 6, 9, 11 and 13 with respect to τ and σ, particularly for Case II.

Some conclusions can be drawn with respect to the type of constraint being considered. (I) Domain constraints (*IC-1*): the test generated will always have $A_{T_i} < A_{IC_i}$ regardless of the fragmentation strategy used. In fact $A_{T_i} = 0$, $\sigma = 1$ and $\tau \approx 0$. This is because the test can be evaluated independently of the rest of the database as it only refers to the tuples to be updated (Ibrahim *et al.* 1998). (II) Key constraints (*IC-2, IC-3*): we have considered: (i) when the global relations involved in the initial constraint are fragmented on the join attribute (*IC-3*), and (ii) when the global relations involved in the initial constraint are not fragmented on the join attribute (*IC-2*). As shown in Tables 4 and 5, it is always possible to derive local tests (i.e. $\sigma = 1$) which are either (a) complete tests for case (i) (tests 4 and 5 of Table 4) or sufficient tests for case (ii) (test 3 of Table 4). These complete tests are cheaper than the initial constraint with respect to the amount of data accessed, i.e. $A_{T_i} < A_{IC_i}$, σ and τ. (III) Referential constraints (*IC-4*): we have considered: (i) when the global relations involved in the initial constraint are fragmented on the reference attribute, and (ii) when the global relations involved in the initial constraint are not fragmented on the reference attribute. For both cases two types of test are generated, namely complete tests (tests 6 and 11, respectively) and sufficient tests (tests 7 and 12, respectively). The sufficient tests are cheaper than the complete tests with respect to σ and τ. (IV) General semantic integrity constraints (*IC-5, IC-6*): In general, $A_{T_i} < A_{IC_i}$ and it is always possible to generate complete tests which are normally global tests, and sometimes possible to generate tests whose $\sigma = 1$ since this depends on the fragmentation rules and the allocation schemas used, also on the complexity of the constraint itself.

In our method, given a set of possible simplified forms (integrity tests), each of the simplified forms is evaluated with respect to the three main components listed above. The most efficient one is selected based on the following heuristic rules (H1–H5). A heuristic based approach is adopted due to: (i) the difficulty in determining all the appropriate components/parameters that can be used to measure and further select the efficient form from a range of possible simplified forms; (ii) the difficulty in assigning suitable weights to all the cost components, where the interaction between these components is not clearly defined; (iii) most of the values assigned to the components/parameters are estimates and not the actual values, which depend solely on the user input (Qian 1989); and (iv) components which are considered as critical factor that can influence the performance of a system in one situation might not be critical in other situations – for example, in a distributed database which is fully replicated, the amount of data transferred across the network in the event of evaluating constraints is not as important as in a distributed database with no replication.

Based on the above arguments, the following heuristic rules are applied:

H1 - Given a set of simplified forms, a simplified form with the lowest σ, A and τ is always preferable to the others.

H2 - As the cost of accessing remote data for verifying the consistency of a database state is the most critical factor that influences the performance of a distributed database, simplified forms which can be evaluated at a single site without involving any interaction with the remote sites are more efficient (Gupta and Widom 1993; Mazumdar 1993). Therefore, a simplified form whose $\sigma = 1$ is always preferable to the others.

H3 - In a situation where more than one local simplified form can be derived, then the simplified form with the lowest A is preferable to the others.

H4 - In a situation where the simplified forms are non-local, then the simplified form with the lowest σ is preferable to the others.

H5 - In a situation where the simplified forms are non-local with the same σ value, then the simplified form with the lowest τ is preferable to the others.

CONCLUSION

In a distributed database, the cost of accessing remote data for verifying the consistency of the database is the most critical factor that influences the performance of the system (Gupta and Widom 1993; Mazumdar 1993; Qian 1989; Simon and Valduriez 1986). In such an environment, simplified constraint forms which can be evaluated at a single site are preferable, i.e. $\sigma = 1$ and $\tau = 0$. In this paper, we have outlined several techniques which are essential for efficient constraint checking of fragmented relations in a distributed database. These techniques, which utilize knowledge about the database application to derive fragment constraints/simplified forms, reduce the amount of data that has to be accessed, the amount of data transferred across the network and the scatter metric which captures the scale of constraint non-locality.

Although many approaches/methods have been proposed for constructing efficient integrity tests from a given integrity constraint and its relevant update operation, these approaches/methods are mostly designed for a centralized environment. Hence, the integrity tests derived by these methods are not suitable for a distributed environment as they often span multiple sites and involve the transfer of data across the network. In this paper, we have shown that it is always possible to generate local tests for domain, key and referential integrity constraints and it is sometimes possible to generate local tests for general semantic integrity constraints, depending on their complexity. These local tests whose evaluation only involves a single site, i.e. the site where the update is to be performed, reduce the amount of data being transferred across the network to perform integrity checks and so improve the efficiency of the integrity checking process. These local tests are derived by employing the integrity test generation techniques.

The overhead of constraint checking in a distributed environment is strongly influenced by the complexity of the integrity constraints, the fragmentation strategies used and the allocation of the fragment relations involved. In this paper, we have focused on four types of constraint, namely: domain, key, referential and simple general semantic integrity constraints. We have analyzed and demonstrated the effect of different fragmentation and allocation strategies on each of these types of constraint and shown that in most cases we gain efficiency.

There are a number of extensions and improvements that could be made: (i) Consider a broader range of constraint types. We have concentrated here on four types of constraint, namely: domain constraints, key constraints, referential integrity constraints and simple general semantic integrity constraints. Other types of constraint, such as aggregate constraints and transition constraints, are worth investigating. (ii) The overhead of constraint checking is also strongly influenced by the type of fragmentation and allocation used. Further investigation of the effect of the fragmentation and allocation strategy on the derived fragment constraints/simplified forms would be worthwhile.

REFERENCES

BARBARA, D. and H. GARCIA-MOLINA. 1992. The Demarcation Protocol: A Technique for Maintaining Linear Arithmetic Constraints in Distributed Database Systems. In *Proceedings of the Conference on Extending Database Technology* (EDBT'92). p. 373-388.

BERNSTEIN, P.A. and B.T. BLAUSTEIN. 1981. A Simplification Algorithm for Integrity Assertions and Concrete Views. In *Proceedings of the 5th International Computer Software and Applications Conference* (COMPSAC'81). p. 90-99.

DATE, C.J. 1995. *An Introduction to Database Systems.* 6th Edition. Addison-Wesley Publishing Company, Inc.

GREFEN, P.W.J. 1993. Combining Theory and Practice in Integrity Control: A Declarative Approach to the Specification of a Transaction Modification Subsystem. In *Proceedings of the 19th International Conference on Very Large Data Bases* (VLDB 19). p. 581-591.

GUPTA, A. and J. WIDOM. 1993. Local Verification of Global Integrity Constraints in Distributed Databases. In *Proceedings of the 1993 ACM SIGMOD Conference.* p. 49-58.

HENSCHEN, L.J., W.W. MCCUNE and S.A. NAQVI. 1984. Compiling Constraint-Checking Programs from First-Order Formulas. *Advances in Database Theory*. Vol. 2. Gallaire, H., Minker, J. and Nicholas, J.M. (Eds). p. 145-169. USA, New York: Plenum Press.

HSU, A. and T. IMIELINSKIN. 1985. Integrity Checking for Multiple Updates. In *Proceedings of the 1985 ACM SIGMOD International Conference on the Management of Data*. p. 152-168.

IBRAHIM, H., W.A. GRAY and N.J. FIDDIAN. 1998. Rule-Based Integrity Enforcement Strategy for a Distributed Database. In *Proceedings of the Third Biennial World Conference on Integrated Design and Process Technology – Issues and Applications of Database Technology* (IADT'98). **2**: 111-118.

IBRAHIM, H., W.A. GRAY and N.J. FIDDIAN. 1996. The Development of a Semantic Integrity Constraint Subsystem for a Distributed Database (SICSDD). In *Proceedings of the 14th British National Conference on Databases* (BNCOD 14). p. 74-91.

MAZUMDAR, S. 1993. Optimizing Distributed Integrity Constraints. In *Proceedings of the 3rd International Symposium on Database Systems for Advanced Applications*. **4**: 327-334.

MCCAROLL, N.F. 1995. Semantic integrity enforcement in parallel database machine. Ph.D Thesis. Department of Computer Science. University of Sheffield. Sheffield (UK).

MCCUNE, W.W. and L.J. HENSCHEN. 1989. Maintaining state constraints in relational databases: a proof theoretic basis. *Journal of the Association for Computing Machinery* **36(1)**: 46-68.

NICHOLAS, J.M. 1982. Logic for improving integrity checking in relational databases. *Acta Informatica* **8(3)**: 227-253.

OZSU, M.T. and P. VALDURIEZ. 1991. *Principles of Distributed Database Systems*. Prentice-Hall International Editions.

QIAN, X. 1989. Distribution Design of Integrity Constraints. In *Proceedings of the 2nd International Conference on Expert Database Systems*. p. 205-226.

QIAN, X. 1988. An Effective Method for Integrity Constraint Simplification. In *Proceedings of the 4th International Conference on Data Engineering* (ICDE 88). p. 338-345.

SIMON, E. and P. VALDURIEZ. 1986. Integrity Control in Distributed Database Systems. In *Proceedings of the 19th Hawaii International Conference on System Sciences*. p. 622-632.

Pertanika J. Sci. & Technol. **10(1)**: 55 - 62 (2002)
ISSN: 0128-7680

Likelihood Based Estimation in the Logistic Model with Time Censored Data

Ayman S. Baklizi
Department of Statistics
Yarmouk University
Irbid-Jordan

Received: 23 October 2000

ABSTRAK

Prosedur-prosedur kesimpulan berdasarkan fungsi kemungkinan dipertimbangkan untuk satu pengagihan logistik dengan data tapisan masa. Pelaksanaan sampel penganggar kemungkinan maksimum yang terhad sepertimana prosedur kesimpulan kemungkinan sampel besar berdasarkan sampel Wald, Rao dan statistik nisbah kemungkinan disiasat. Hasilnya, apa yang diperolehi daripada pengagihan normal asimptot penganggar kemungkinan maksimum didapati tidak tepat. Hasil siasatan juga menunjukkan penganggaran jeda yang berdasarkan statistik Wald dan Rao memerlukan lebih banyak saiz sampel berbanding penganggaran jeda berdasarkan statistik nisbah kemungkinan untuk memperolehi ketepatan yang munasabah.

ABSTRACT

Inference procedures based on the likelihood function are considered for the one logistic distribution with time censored data. The finite sample performances of the maximum likelihood estimator as well as the large sample likelihood inferential procedures based on the Wald, the Rao, and the likelihood ratio statistics are investigated. It is found that the obtained from the asymptotic normal distribution of the maximum likelihood estimator are found no accurate. It is found also that interval estimation based on the Wald and Rao statistics need much more sample size than interval estimation based on the likelihood ratio statistics to attain reasonable accuracy.

Keywords: Confidence intervals, logistic distribution, maximum likelihood estimator, time censored data

INTRODUCTION

The logistic distribution arises in a variety of fields, for example it can be applied as a growth model in human populations and certain biological organisms (Pearl *et al.* 1940). Oliver (1964) used this distribution to model agricultural production data. It also arises in the analysis of survival data (Placket 1959) as well as the analysis of income distributions (Fisk 1961). Many other applications and motivations for this distribution are discussed in Balakrishnan (1992) and Johnson *et al.* (1994).

Inference procedures for the parameters of this distribution have been discussed in the literature. Harter and Moore (1967) and Balakrishnan (1992) discussed and investigated the properties the maximum likelihood estimator in sample of size 10 and 20, for various choices of type 2 censoring. Confidence intervals for the parameters of this models are constructed by Antle *et al.* (1970) by simulating the percentage points of some pivotal quantities based on the maximum likelihood estimator. Schaffer and Sheffield (1973) have further discussion on this problem while Bain *et al.* (1992) considered this problem when the data is type 2 censored.

In this paper I shall consider maximum likelihood estimation and asymptotic interval estimators based on inverting the likelihood based statistics when the data is time censored. These statistics are the Wald, the Rao, and the likelihood ratio (Barndorff- Nielsen and Cox 1994). These statistics are often used for interval estimation with censored data (Nelson 1990), and are known to have an asymptotic chi-squared distribution (Rao 1973). However their performances in finite samples are different and change from one model to another (Lawless 1982). Hence it is desirable to use the statistics which has a faster covergence rate to its limiting distribution and therefore is applicable for small sample sizes that may be in practice (Cox 1988).

METHOD

The Model and the Likelihood Statistics

The probability density function and the cummulative distribution function of the logistic distribution are given respectively by (Johnson *et al.* 1994)

$$f(x,\mu,\sigma)=\frac{c}{\sigma}\frac{\exp(-\frac{c(x-\mu)}{\sigma}}{\left(1+\exp(-\frac{c(x-\mu)}{\sigma}\right)^{2}}\qquad x>0,\ \sigma>0,\quad -\infty<\mu<\infty$$

where $c=\frac{\pi}{\sqrt{3}}$. Consider a random sample of size N from this distribution, of which n are less than or equal to t (some predetermined censoring time) and the remaining no observations are censored and are only known to exceed t. Thus $N = n + n_0$. Here t is fixed but n and n_0 are both random the likelihood function is given by

$$L(x,\mu,\sigma)=\left(\frac{c}{\sigma}\right)^{n}\exp\left(-\sum_{i=1}^{n}\frac{c(x_i-\mu}{\sigma}\right)\prod_{i=1}^{n}1+\exp(-\frac{c(x_i-\mu)}{\sigma}^{-2}\times$$

$$\prod_{i=1}^{n_0}\left(1-\left(1+\exp(-\frac{c(t-\mu)}{\sigma}\right)^{-1}\right)$$

The maximum likelihood estimator can be found by solving the system of first partial derivatives of the log-likelihood function, describe in the appendix. To present the Wald statistics, the Rao and the likelihood ratio statistics for this model we need the following quantities.

$$U(\mu,\sigma) = \begin{pmatrix} \dfrac{\partial l(\mu,\sigma)}{\partial\mu} \\ \dfrac{\partial l(\mu,\sigma)}{\partial\sigma} \end{pmatrix} \text{and} \quad I(\mu,\sigma) = \begin{pmatrix} \dfrac{-\partial^2 l(\mu,\sigma)}{\partial\mu^2} & \dfrac{-\partial^2 l(\mu,\sigma)}{\partial\mu\partial\sigma} \\ \dfrac{-\partial^2 l(\mu,\sigma)}{\partial\mu\partial\sigma} & \dfrac{-\partial^2 l(\mu,\sigma)}{\partial\sigma^2} \end{pmatrix}$$

The Wald statistics for μ and σ are given respectively by

$$m_1(\mu) = (\hat{\mu} - \mu)^2 (I^{11}(\mu,\tilde{\sigma}))^{-1}$$
$$S_1(\sigma) = (\hat{\sigma} - \sigma)^2 (I^{22}(\hat{\mu},\hat{\sigma}))^{-1}$$

The Rao statistics for μ and σ are given respectively by

$$m_2(\mu) = (U_1(\mu,\tilde{\sigma}))^2 I^{11}(\mu,\tilde{\sigma}),$$
$$S_2(\sigma) = (U_2(\tilde{\mu},\sigma))^2 I^{22}(\tilde{\mu},\sigma)$$

The likelihood ratio statistics for μ and σ are given respectively by

$$m_3(\mu) = 2(l(\hat{\mu},\hat{\sigma}) - l(\mu,\tilde{\sigma}),$$
$$S_3(\sigma) = 2(l(\hat{\mu},\hat{\sigma}) - l(\tilde{\mu},\sigma))$$

Where $l(\mu,\sigma) = \ln(L(\mu,\sigma))$ is the log-likelihood function, $\hat{\mu}$ and $\hat{\sigma}$ are the maximum likelihood estimator of μ and σ. $\tilde{\mu}$ is the maximum likelihood estimator of μ for a given value of σ. I^{ij} is the *ij*-th element of I^{1}th element of I-1, the inverse of the information matrix.

Finite Sample Performance of the Likelihood Statistics

In this section we shall describe a simulation study conducted to investigate the finite sample behaviour of the maximum likelihood estimator, and confidence intervals based on the Wald, the Rao, and the likelihood ratio statistics.

The criteria used for the evaluation of the performance of the maximum likelihood estimator are the bias, the finite sample variance, and the adequacy of the asymptotic variance estimates (Elperin and Gertsbakh 1987). For the confidence intercals, we use the attainment of the nominal error probabilities and the symmetry of lower and upper error probabilities (Jennings 1987). Attainment of nominal error probabilities and our conclusions therefore are imprecise and can be misleading. The symmetry of lower and upper error probabilities means that of the interval fails to contain the true value of the parameter, it is equally expect this symmetry because they are using symmeteric percentiles of the approximating distributions to form their confidence intervals.

TABLE 1
Values of the bias, finite sample variance, mean squared error and asymptotic variance for the maximum likelihood estimator

		Location Parameter				Scale Parameter			
CP	N	Bias	FSV	MSE	ASV	Bias	FSV	MSE	ASV
0.0	2.0	-59	436	437	449	-308	345	354	339
	40	-40	219	219	225	176	163	166	171
	60	-27	152	152	151	-113	115	116	115
	80	6	110	110	113	-80	82	82	87
	100	8	88	88	91	-77	68	68	69
	120	11	76	76	76	-69	58	59	58
	140	13	64	64	65	-60	49	50	50
0.1	60	-26	152	152	152	-94	126	127	128
	80	6	110	110	114	-71	89	90	96
	100	9	87	87	91	-69	74	75	77
	120	11	76	76	76	-65	63	63	64
	140	13	64	64	65	-58	53	54	55
0.3	60	-13	157	157	160	-69	170	171	175
	80	23	114	114	120	-35	129	129	132
	100	21	90	90	96	-42	103	103	104
	120	23	77	78	79	-38	86	86	87
	140	24	66	66	68	-33	73	74	74
0.5	60	22	203	203	208	-41	269	269	274
	80	53	143	144	154	-5	197	197	205
	100	56	116	117	122	-2	162	162	163
	120	57	103	104	101	0	133	133	135
	140	56	87	88	86	4	113	113	115

However, symmetry of error probabilities may not occur due to the skewness of the actual sampling distribution (Jennings 1987).

In the simulations we use values of the sample size N as 60, 80, 100, 120 and 140. In each case we examined the censoring proportions 0.1, 0.3, and 0.5. For every combination of the sample size and censoring proportion we generated 2000 samples which are used to determine, the bias, the finite sample variance, the asymptotic variance of the maximum likelihood estimator, the mean squared errors of the estimators (MSE), and the error probabilities of the confidence intervals obtained from the Wald the Rao, and the likelihood ratio statistics. The levels of significance used are α= 0.05, and 0.1. The results are given in Tables 1, 2 and 3. All values in the tables are multiplied by 10000.

FINDINGS AND CONCLUSIONS

Concerning the behaviour of the maximum likelihood estimator for the location and scale parameters, it appears that it is almost unbiased. It appears

TABLE 2
Lower and upper error probabilities of confidence intervals based on the Wald, Rao and likelihood ratio statistics. a = 0.1

		Location Parameter						Scale Parameter					
		Wald		Rao		LR		Wald		Rao		LR	
CP	N	L	U	L	U	L	U	L	U	L	U	L	U
0.0	20	300	235	100	95	255	195	995	20	0	525	435	115
	40	350	225	195	130	315	190	730	30	0	445	340	115
	60	310	290	225	200	285	265	590	50	0	430	345	140
	80	225	210	180	185	210	210	490	70	20	340	315	175
	100	225	225	190	195	220	220	500	65	55	375	335	155
	120	260	250	210	220	250	240	515	85	80	380	365	155
	140	275	275	240	235	265	260	480	110	55	395	340	180
0.1	60	310	275	215	190	290	255	585	40	0	490	310	155
	80	235	210	175	185	215	210	475	55	15	385	300	140
	100	220	225	175	200	215	220	540	65	40	390	325	140
	120	250	250	210	210	240	250	505	55	65	375	37	165
	140	290	275	240	240	270	265	450	90	55	380	340	175
0.3	60	295	210	185	255	260	225	570	25	0	570	330	185
	80	260	195	140	220	235	210	470	65	10	550	290	230
	100	260	220	165	255	235	235	510	85	20	445	310	225
	120	280	220	210	260	255	230	485	75	35	495	330	185
	140	295	220	220	270	275	245	475	120	30	485	265	230
0.5	60	365	85	45	520	295	275	610	40	0	615	295	195
	80	300	115	60	435	220	240	545	50	0	610	250	190
	100	295	110	65	425	225	215	595	90	0	560	275	235
	120	360	165	105	470	255	250	510	105	0	500	300	210
	140	340	165	105	460	235	265	460	105	0	530	270	210

also that the bias decreases when increasing the sample size, as anticipated from the asymptotic unbiasedness of maximum likelihood estimators. The bias and variance increase when increasing the censoring proportion. The variance of the maximum likelihood estimator decreases as the sample size increases, this is because of its consistency (Rao 1973). The asymptotic approximation to the variance of the maximum likelihood estimator provided by the observed information matrix seems to hold very well in this model, even for small samples with high censoring level. This also shows the high efficiency of the maximum likelihood estimator in this model.

For the interval estimation of the location parameter, it is clear that intervals based on the Wald statistics perform well. However, as the censoring proportion increases they tend to be asymmetric, especially when the sample size is small. Intervals based in the Rao statistics are generally conservative, that is, having an actual coverage probability that is less than the nominal. Moreover they tend

TABLE 3
Lower and upper error probabilities of confidence intervals based on the Wald, Rao and likelihood ratio statistics. a = 0.1

		Location Parameter						Scale Parameter					
		Wald		Rao		LR		Wald		Rao		LR	
CP	N	L	U	L	U	L	U	L	U	L	U	L	U
0.0	20	555	555	340	285	505	515	1340	100	0	735	865	310
	40	615	535	470	385	590	500	1055	115	25	635	780	270
	60	590	505	520	450	570	495	980	180	135	700	655	330
	80	480	445	445	425	475	435	800	230	180	565	580	315
	100	470	505	430	460	460	495	795	230	240	655	640	365
	120	535	490	490	460	525	475	810	275	325	640	675	380
	140	505	500	470	450	500	495	785	280	270	625	625	425
0.1	60	570	510	495	460	570	505	930	165	110	705	660	385
	80	495	440	430	430	455	440	770	190	150	605	575	340
	100	475	520	415	445	470	490	820	200	215	690	66-	375
	120	525	490	475	455	520	480	860	235	290	650	650	360
	140	490	495	460	480	485	495	775	235	270	665	635	390
0.3	60	605	440	375	515	515	470	930	170	40	820	605	400
	80	505	370	375	465	475	420	830	255	105	745	560	425
	100	465	455	405	520	445	485	820	245	160	725	590	410
	120	515	435	430	500	500	455	815	240	165	7-5	630	440
	140	535	490	475	540	525	510	785	315	155	735	580	450
0.5	60	705	300	290	740	570	460	935	145	0	955	565	510
	80	520	285	235	715	435	415	820	165	10	860	555	465
	100	545	280	255	740	470	455	885	250	40	840	63	410
	120	615	370	335	730	520	500	820	255	80	700	585	380
	140	615	385	320	745	495	530	785	285	90	790	540	440

to be asymmetric as the censoring proportion increases. On the other hand, interval based on likelihood ratio statistics tend to attain their nominal coverage probability and are symmetric in almost all situations considered.

For the scale parameter, it is clear that intervals based on the Wald and the Rao statistics are highly asymmetric, even for samples as large as 140, and hence cannot be recommended for use in practice. While intervals based on the likelihood ratio statistics still attain their nominal coverage probability and are symmetric in almost all situations considered.

The high asymmetry of Wald and Rao intervals indicates that the actual sampling distributions of the Wald and the Rao statistics are highly skewed (Jennings 1986), and they require a large sample size for the asymptotic chi-squared approximation to hold well. While the likelihood ratio statistics appears to need much lesser sample size to justify its use.

This shows that the likelihood ratio statistics converges of its limiting distribution much faster than the Wald and the Rao statistics.

The likelihood ratio statistics are applicable for one sided interval estimation and for one sided hypotheses testing because of their symmetric lower and upper error probabilities. This is, however, not the case for Wald and Rao statistics unless the sample size is high.

For all kinds of intervals considered it appears that as the sample size increases, all intervals tend to have error probabilities that are more symmetric and closer to the nominal ones. Also, larger nominal error probabilities are attained faster than smaller error probabillities.

ACKNOWLEDGEMENT

The author is grateful to the referee for his comments which have improved the paper.

REFERENCES

ANTLE, C., KLIMKO and W. HARKNESS. 1970. Confidence intervals for the parameters of the logistic distribution. *Biometrica* **57**:397-402.

BAIN, L. J. EASTMAN, M. ENGELHARDT, N. BALAKRISHNAN and C. ANTLE. 1922. Reliability Estimation Based on MLEs for Complete and Censored Samples. In *Handbook of the Logistic Distribution,* ed. N. Balakrishznan. New York: Marcel Dekker.

BALAKRISHNAN, N. 1922. Handbook of the Logistic Distribution. In *Inference and Asymptotics,* ed. Barndorff-Nielsen, O. and D.R. Cox. 1994. New York: Marcel Dekker. Chapman and Hall.

COX, D. R. 1988. Some aspects of conditional and asymptotic inference: a review. *Sankhya. A.* **3**: 314-337.

ELPERIN, T. and I. GERTSBAKH. 1987. Maximum likelihood estimation in a Weilbull regression model with type 1 censored data: a Monte Carlo study. *Comm. Stat. Simul. Comput.* **16(2)**: 349-371.

FISK, P. 1961. The graduation of income distributions. *Econometrika* **29**: 171-185.

JENNINGS, D. 1986. Judging inference adequacy in logistic regression. *JASA* **81(394)**: 471-476.

JENNINGS, D. 1987. How do we judge confidence intervals adequacy? *The American Statistician* **41(40)**: 335-337.

JOHNSON, N., S. KOTZ and N. BALAKRISHNAN. 1994. *Continuous Univariate Distributions.* Volume 1. New York: Wiley.

HARTER, H. and A. MOORE. 1967. Maximum likelihood estimation from censored samples of the parameters of the logistic distributions. *J. Amer. Stat. Assoc.* **62**:657-684.

LAWLESS, J. 1982. *StatiModels and Methods for Ligetime Data.* New York: Wiley.

OLIVER, F. 1964. Methods of estimating the logistic growth function. *Appl. Statist.* **13:** 57-66.

PEARL, R., L. REED and J. KISH. 1940. The logistic curve and the census count of 1940. *Science* **92:** 486-488.

PLACKETT, R. 1959. The analysis of Life Test Data. *Technometrics* **I:** 9-19.

RAO, C. 1973. *Linear Statistical Inference and Its Applications.* New York: Wiley.

Schafer, R. and T. Sheffield. 1973. Inferences on the parameters of the logistic distribution. *Biometrika* **29:** 449-455.

Pertanika J. Sci. & Technol. **10(1)**: 63 - 71 (2002)
ISSN: 0128-7680

Operational Characteristics and Determination of Resistance for Effective Powering and Propulsion of Fishing Boats of Lower Perak River of Malaysia

M. A. Hossain Mollah, S.M. Sapuan, M. Megat Ahmad, Z. Leman & M.Y. Ismail

Department of Mechanical and Manufacturing Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor

Received: 17 August 1999

ABSTRAK

Kertas kerja ini melaporkan beberapa kajian kes terhadap ciri-ciri rintangan dan kejangan tipikal bot-bot kecil yang digunakan oleh nelayan di hilir Sungai Perak, Malaysia. Matlamat kajian ini ialah untuk menyediakan satu kaedah yang lebih baik dan efisyen bagi menentukan rintangan dan kejangan bagi kuasa dan dorongan yang berkesan bagi bot kecil. Anggaran terhadap ciri-ciri kipas enjin yang sesuai telah dilakukan menggunakan rajah siri SK. Kecekapan proses penukaran kuasa enjin kepada kuasa tujahan yang efektif boleh dipertingkatkan dengan memilih gabungan enjin dan parameter kipas yang sesuai. Kuasa yang diperolehi oleh enjin dalam kajian kes ini didapati berada di antara 6.4 kW hingga 7.4 kW untuk kelajuan kipas 1000 rpm dan 3000 rpm; dan dengan kelajuan bot, Vs = 8.5 knot. Oleh itu nisbah kuasa-anjakan bot berubah daripada 1.28 kW/ton kepada 1.5 kW/ton; bergantung kepada kelajuan kipas. Usaha telah dijalankan untuk memperolehi kuasa enjin dan dorongan yang sesuai bagi bot-bot kecil dengan tujuan untuk mengurangkan kos penggunaan bahan api.

ABSTRACT

The case study of a resistance and propulsion characteristics of a typical small traditional fishing boats of lower Perak River of Malaysia is reported in this paper. The aim of this study is to provide a better or more efficient method of determining the effective powering and propulsion of fishing boats. For the estimation of propeller characteristics, the propeller "SK" – series diagrams were used. The results of the calculations show that the efficiency of the conversion process of the engine power into the effective thrust can be improved by proper selection of the engine, propeller and its parameters. The power requirement of the engine according to the calculation was found to be 6.4 kW and 7.5 kW for the propeller speeds of 1000 rpm and 300 rpm respectively for Vs = 8.5 knots. Thus the power-displacement ratio varies from 1.28 kW/ton depending on the propeller speed. Efforts have been made for proper powering and propulsion of the traditional fishing boats in order to minimize the fuel cost.

Keywords: Fishing boats, fishing gear, propeller characteristics, resistance and powering

INTRODUCTION

The fishing sector plays an important role to national economy. It is most frequently measured by the contribution to GNP, the number of fishermen employed and the nutritional aspects such as annual per-capita of fish consumption and the percentage of protein intake contribution by fisheries products. Fisheries sector contributes approximately 2 % of the gross domestic product of the Peninsular Malaysia. In 1996 the total production from fisheries sector amounted to 1,239,434 tones valued at RM 3.84 billion, while in 1973, the quantity of fish caught was 365,000 tons valued at Ringgit Malaysia, RM 0.884 billion (Department of Fisheries, Malaysia 1996). The aim of this study is to provide a better or more efficient method of determining the effective powering and propulsion of fishing boats. This paper is restricted to consider the traditional fishing boats of the Lower Perak River of Malaysia.

Characteristics of Traditional Fishing Boats

Normally, wooden boats with a variety of size and shape are used for fishing in almost all over the world. However, there is a recent trend in using polymeric-based composite material in building small and medium sized boats Sapura *et al.* (1999). Most often, the boats are grouped into the vessel size, power unit, type of gear and the operating distance from shore. For example, Indonesia classified the boats based on size and whether or not the vessel is mechanized. Thailand's distinction between small-scale and large-scale is based on the type of fishing gear used. In the Philippines, all fishermen using vessels over 3 tons are considered commercial whereas the fishermen using vessels of less than 3 tons are considered municipal fishermen. Hong Kong and Singapore classified the fishing boats as inshore and offshore fisheries. Malaysia takes into account vessel displacement, type of gear used, and area fished (Smith 1983).

Traditional fisheries are carried out in small-scale fishing units. They often consists of kin groups using small and occasionally powered boats. The fishing activity is often part-time, and household income may be supplemented by other activities of the fishermen. Gear, such as nylon netting, which may be made by machines, is usually operator-assembled. It requires minimal or no mechanical assistance to operate. Investment levels of the fishing implements are low. Catch per fishing unit and productivity per fisherman range from medium to very low. Traditional fishing communities are frequently isolated, both geographically and socially, and the standard of living of traditional fishing households is low to minimal.

The major problems that contribute to low income and low standards of living are limited fisheries resources, inadequate or improper uses of fishing implements (boats, engines, propeller, fishing nets etc.) and lack of market power.

Fishing Boats of Lower Perak River

A variety of fishing boats was observed during the case study in the lower Perak River of Malaysia. The fleet strength of the region is approximately 200.

The boats are usually made of locally available wood by local carpenters with their own individual initiative. The construction methods are relatively simple and skills are locally available. The total price of the boat with propulsion accessories is about RM 3000. The main technical drawbacks of implements are that they are based on non-standard design. The cheaper second-hand automobile engines, usually with the gearboxes, are installed. The design of the propeller is not based on standard criteria, the diameter, pitch, efficiency of the propeller are not at all considered. Consequently, each boat is not provided with matching and properly designed propeller resulting in uneconomical utilization of the engine power.

The purpose of study is to find out the power requirement and the optimum propeller parameters to improve the fuel efficiency for the operation of the fishing boats. Analysis is performed on the basis of the hull specification of a wooden boat, which is commonly found in the lower Perak River of Malaysia as shown in *Fig. 1*.

Fig. 1: A typical fishing boat in lower Perak River

a) Hull Specification

The following are the specifications of the hull of a selected fishing boat studied as shown in *Fig. 1:*
Length overall, m L_{OA} = 12.4
Length at waterline, m L_{WL} = 11.0
Breadth at waterline, m B_{WL} = 1.8
Draught, m T = 0.46

Displacement, tons D = 5.0
Speed, knots Vs = 7.5
Hull Form/form coefficient, ~ 0.55 (Anon 1990)
Fishing mode: Trawling net
Boat Class: Traditional rural boat builders (no classification)
The installed Engine of the boat: 3-cylinder Yanmar 25G, Rated Power 20 kW

b) Engine Specifications

During the case study, the authors observed a variety of second-hand, 2 to 4 cylinder engines in similar boat sizes. The rated power of engines was found from 10 kW to 35kW. Thus, boats were found in operation in a wide range of power-displacement ratio (from 2.o to 7). The similar situation was also observed by the other researchers (Overa and Ravikumar 1989) for the traditional fishing boats of Bay of Bengal. In most cases the boats were incorporated with the gearboxes, with the reduction ratio of 3:1. Propeller speed generally ranged from 800 to 1200 rpm. The authors also observed the different types of coupling arrangements from gearbox to propeller shaft. The most commonly used engine models are 3-cylinder Yanmar 25G and 4-cylinder Mazda.

Resistance and Powering

The power required to drive a fishing boat depends upon the resistance offered by the water, air, trawling gear, propeller efficiency, hull shape of the boat and also on means of transmission. The resistance of a boat depends upon multitude of factors such as geometrical configuration of the boat, its speed, properties of water etc. (Mollah and Sarkar 1990; Mollah 1983). Because of the complexity of the problem, it is very difficult to predict the resistance of a small fishing boat accurately. In general, there is no simple solution to this problem. If the resistance and propulsion tests are made with a model of the standard boat, it will be easier to predict the required power for propulsion with sufficient accuracy. However, this type of tests can only be performed in a well-equipped laboratory and may not be economically viable.

Many years of experience of authors shows that estimation of resistance by using the ITTC formulae and by Pampel method give quite satisfactory results (Mollah 1990; Kulagin 1974). Moreover, the total resistance, RT was analyzed for small boats (Mollah and Sarkar 1990) and it was reported that the total resistance obtained by both methods was almost the same up to the boat speed of 6 knots. But at higher speed the value of resistance obtained by the ITTC formulae was 3-5% more than the value obtained by the Pampel method.

This paper considers both methods to predict the resistance of the boat and for powering calculations the higher resistance value was taken to have more power margins for more adverse conditions of the fishing boats.

The total resistance, RT is calculated from the ITTC formulae after estimating the wetted surface area (Kulagin 1974). The wetted surface area was obtained by the following equation:

$$S = [2\,T + 1.37\,(C_b - 0.274)\,B_{WL}]\,L_{WL} \tag{1}$$

The value of C_b is 0.55 (Anon 1990) and

$$R_T = C_T \rho \frac{V^2}{2} S \tag{2}$$

C_T is calculated according to the methodology of Kulagin (1974).

According to the Pampel method, the required power to overcome the resistance and the total resistance is calculated by the following equations:

$$N_R = 0.735 \frac{W\ V_S^3\ K}{L_{WL}\ C_P\ \lambda} \sqrt{\Psi_1} \tag{3}$$

K is coefficient that considers the extended parts of the boat and the value of K is taken as 1.00 for a single propeller (Kulagin 1974). Similarly λ and ψ_1 were calculated according to the methodology described by Kulagin (1974).

and

$$R_T = \frac{N_R}{V} \tag{4}$$

The total resistance was estimated by above mentioned methods and subsequently the required power, N_E and N_R were also estimated. N_E is the required power to overcome the resistance of the boat and N_R is the required engine power to drive the boat effectively. It may be mentioned here that such calculations of course assume calm water conditions and power margins would need to be allowed for more adverse conditions. Therefore, the prime-mover has to provide 20% extra power margin than is required to overcome the resistance considering the propulsive efficiency. Moreover, the fishing boats are practically always required to operate in conditions which are different from the calm water conditions and the fishing boats often have the power requirements dictated by the trawl gear drag and speed of the trawl. The trawl gears and the trawling speed offer an additional 20% resistance to the boat (Fyson 1985).Thus, a prime mover of fishing boats require a total 40% extra power margin from the calculated one.

Propeller Parameters

The main purpose of the propeller is to convert the engine power into propulsive thrust. The rotational speed of the propeller is the most important factor to increase the efficiency of conversion process of the engine power into thrust. In general, the slower the speed of the propeller, the more efficient in converting the engine power into thrust. But the speed of the engine is too high for a direct drive. Therefore, a gearbox with suitable reduction ratio is one obvious solution for optimum utilization of the engine power and properly selected propeller parameters can improve fuel economy and speed. The larger diameter propeller is more efficient, but not always suitable for installations in small boats due to the shallow draught. Therefore it is better to select a propeller with the maximum possible propeller pitch. The propeller pitch is related to the power available, the reduction ratio used in the gearbox, and the speed of the engine. These parameters have to be assumed together to arrive at a satisfactory combination and to determine the diameter and pitch of the propeller. On the basis of these calculations the optimum propeller parameters were determined. A rough rule for the propeller diameter must be less than two-third of the draught aft (Kulagin 1974).

$$D_{\max} = \frac{2}{3} T_A \tag{5}$$

The propeller diameter and its pitch ratio H/D can be estimated as first approximation by the following equations:

$$D = \frac{\chi}{\sqrt{n_p}} \sqrt{\frac{N_R}{V_s}} \tag{6}$$

The value of coefficient χ depends on the disc ratio and number of blades of the propeller and was determined by using the table 16.2 (Kulagin 1974)

$$\frac{H}{D} = 1.65 \left(1 - 0.11 \sqrt{n_p} \sqrt{\frac{N_R}{V_s^5}} \right)^2 \tag{7}$$

On the basis of this approximate propeller diameter, five values of D were assumed within the range 0.7 – 1.4 times of D (original/initial) and the optimum propeller parameters (Pitch and efficiency) were determined.

For the estimation of propeller characteristics, the propeller "SK"-series diagrams were used (Anon 1990; Voitkunsky *et al.* 1973). The results of the calculations are shown in *Figs. 2* and *3*.

RESULTS AND DISCUSSIONS

Figs. 2 and *3* show the mutual adjustment of power requirements and the propeller characteristics. The curve NR = f (n) is the power absorption at various speeds of the propeller. The power absorption of the propeller increases with the increase in propeller speed. The efficiency of the propeller is less at higher speed due to the excessive circumferancial losses of the propeller.

Fig. 3 shows that the propeller diameter and its efficiency increase with the decrease in propeller speed. The rate of increasing of its parameters is much higher with the decrease in propeller speed at 1000 rpm and lower. Therefore the further reduction of the propeller speed may not be suitable propeller diameter for installation in a small boat due to the shallow draught of the boat. The power requirement of the engine according to the calculation was found to be 6.4 kW and 7.5 kW for the propeller speed of 1000 rpm and 3000 rpm respectively for Vs = 8.5 knots (*Fig. 2*). Thus the power-displacement ratio varies from 1.28 kW/ton to 1.5 kW/ton depending on the propeller speed.

Considering a total of 40% more power margins (Fyson 1985) for more adverse conditions of the fishing boats, the required engine power would be 9.0 kW to 10.5 kW for the propeller speed of 1000 rpm and 3000 rpm respectively. The power prediction approaches of the principal author in his earlier publication agreed with the current result (Mollah 1990). Therefore, a single cylinder or 2-cylinder engine with 10 kW power range may be considered quite enough for the economical operation of the fishing boat. However it may require further study for the clarification of the estimated power requirements for the particular type and size of a boat.

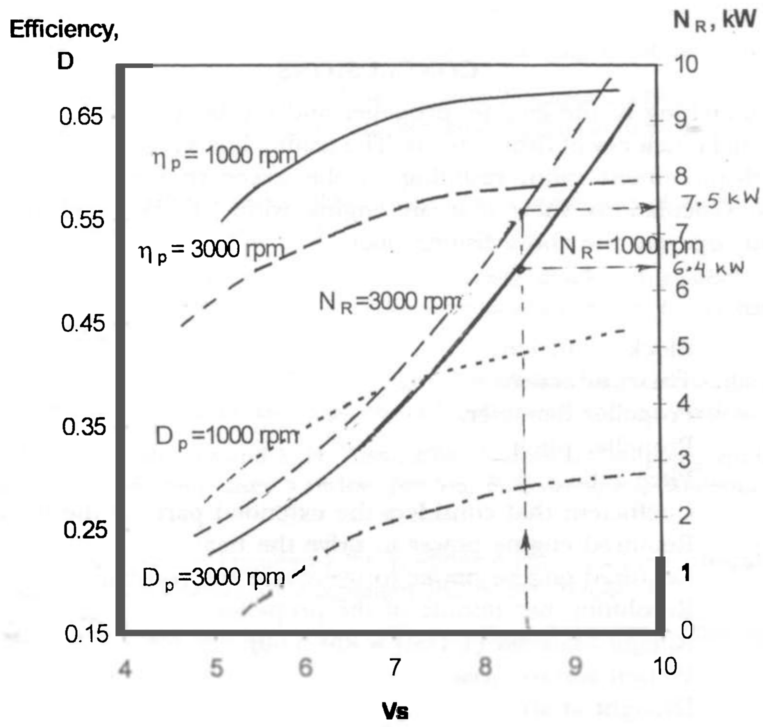


Fig. 2: Mutual characteristics of propeller and boat speed

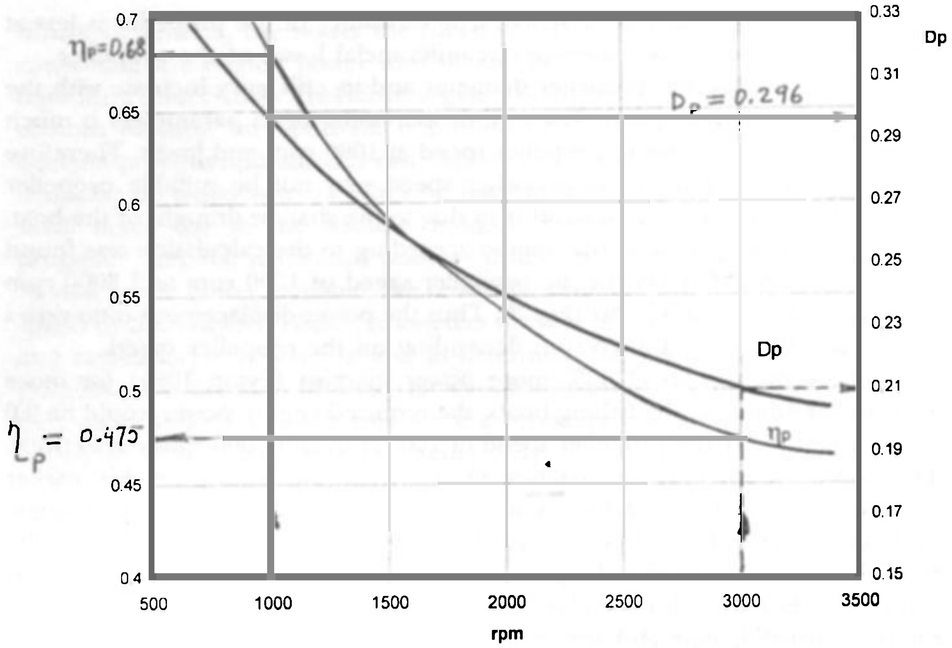


Fig. 3: Mutual characteristics of propeller and engine power

CONCLUSIONS

Proper matching of the engine, propeller and the boat size may improve the operational efficiency of fishing boats. The fishing boats are operating at higher power displacement ratio, resulting in the excessive fuel consumption of engines. Calculations show that an engine with 10 kW rated power may efficiently operate the above fishing boat.

Notations

C_b	Block coefficient
C_T	Total resistance
D	Propeller diameter.
H	Propeller pitch
H/D	Pitch ratio
K	Coefficient that considers the extended parts of the boat
N_R	Required engine power to drive the boat
N_E	Required engine power to overcome the resistance.
n_p	Revolution per minute of the propeller
RM	Ringgit Malaysia (1 USD = RM 3.80)
S	Wetted surface area
T_A	Draught at aft
V	Speed of the boat in m/sec

V_s	Speed of the boat in knots
W	Displacement, ton
χ	Coefficient which considers the length of the boat
ρ	Density of water
ψ	Wake coefficient
η_p	Efficiency of the propeller
λ	Coefficient which considers the overhanging parts

ACKNOWLEDGEMENTS

The authors wish to thank boat owners of the Kampung Hujung Rintis, Kota Setia, Teluk Intan, Perak for their cooperation in taking measurement and providing the required information regarding the fishing boats. The authors are also indebted to the Department of Mechanical and Manufacturing Engineering, UPM for providing the funds for this study.

REFERENCES

ANON. 1990. Yanmar Diesel Engine Instruction Book, Marine, no. 3, Yanmar Diesel Company Ltd., Japan.

DEPARTMENT OF FISHERIES, MALAYSIA. 1996. *Annual Fisheries Statistics,* 1996, Volume 1, Kuala Lumpur: Department of Fisheries Malaysia.

FYSON, J. 1985. *Design of Small Fishing Vessels.* Surrey: Fishing News Books Ltd.

KULAGIN, V.D. 1974. *Theory and Construction of Fishing Boat.* Ship Building, Leningard.

MOLLAH, M.A.H. 1983. Design of Power Installation of Transport Refrigerating Vessels, 9000 Tons Capacity, M. Sc. Engg. Thesis, Astrakhan Technical Institute of Fisheries, USSR.

MOLLAH, M.A.H. and M.A.R. SARKAR. 1990. Powering and propulsion of a country boat. *J. Ins. Eng.* **18(4):** 5-10. Bangladesh.

MOLLAH, M.A.H. 1990. Performance analysis of locally developed gearbox for the country-boats of Bangladesh. *J. Nat. Oceonographics and Maritime Ins. of Bangladesh (NOAMI)* **7(1&2):** 15-20.

OVERA, A. and R. RAVIKUMAR. 1986. Pivoting Engine Installation for Beach-landing Boats, Bay of Bengal Programme, Development of Small-Scale Fisheries, Madras.

SAPUAN, S.M., M.A.H. MOLLAH and M.H. YUNIN. 1999. Marine and shipping applications of polymer–based composites: a review. *Jurutera, Bull. Ins. of Eng. Malaysia* **(7):** 51-55.

SMITH, I.R. 1983. A Research Framework for Traditional Fisheries, International Center for Living Aquatic Resources Management (ICLARM), Manila.

VOITKUNSKY, Y.I, R.Y. PERSIC and I.A. TITOV. 1973. *Handbook of Ship Design.* Ship Building, Leningard.

Pertanika J. Sci. & Technol. **10(1)**: 73 - 88 (2002)

ISSN: 0128-7680

Kajian Perbandingan di antara Ujian H-Ometer dan Ujian Brazilian

Husaini Omar

Mountainous Terrain Development Research Centre,
Fakulti Kejuruteraan, Universiti Putra Malaysia,
43400 UPM Serdang, Selangor

Diterima: 25 September 2000

ABSTRAK

Kajian perbandingan telah dijalankan ke atas ujian H-Ometer dengan ujian yang sedia ada iaitu ujian Brazilian. Kajian perbandingan ini telah dijalankan dengan menggunakan spesimen buatan. Ini adalah kerana spesimen buatan diketahui kandungannya dan ia juga adalah spesimen yang isotropik dan homogen. Oleh itu spesimen adalah seragam. Keputusan kedua-dua ujian dibincangkan. Hasil keputusan menunjukkan kekuatan tegangan dari ujian H-Ometer lebih besar berbanding kekuatan tegangan dari ujian Brazilian. Hasil dari perbandingan antara kedua-dua keputusan didapati kekuatan tegangan dari ujian H-Ometer, σ_{ho}, adalah sama dengan 1.043 kali kekuatan tegangan dari ujian Brazilian, σ_b. Hasil perbandingan ini menunjukkan bahawa kekuatan tegangan dari ujian H-Ometer adalah berkait rapat dengan kekuatan tegangan dari ujian Brazilian. Ini menunjukkan satu keputusan yang amat baik.

ABSTRACT

The comparison study is carried out between H-Ometer test and Brazilian test. In this study, the artificial weak rock or make-up sample is used. This is because the contents of the materials are known, the sample can be controlled and it is assumed to be isotropic and homogenous. Both of the test results are reported and discussed. The tensile strength from the H-Ometer, σ_{ho}, test shows slightly higher than the Brazilian test, σ_b. It is equal to 1.043 times of the tensile strength from the Brazilian test. Based on the correlation, the H-Ometer test shows a good agreement with the Brazilian test.

Kata kunci: Ujian H-Ometer, ujian Brazilian, spesimen buatan

PENGENALAN

H-Ometer adalah alat baru yang direka cipta untuk mengukur kekuatan tegangan ke atas batuan lemah. Alat ini mempunyai saiz yang kecil di mana garis pusatnya adalah 12mm dan panjangnya adalah 74mm. Disebabkan oleh kepayahan yang biasanya dihadapi untuk menjalankan ujian terus tegangan sepaksi ke atas bahan-bahan rapuh, seperti batuan lemah telah menyebabkan banyak kaedah ujian secara tak langsung telah dihasilkan untuk mendapatkan kekuatan tegangan. Ujian 'Brazilian' merupakan yang paling popular. Ianya pertama kali telah diperkenalkan untuk menguji konkrit oleh Carneiro di Brazil dan Akazawa di Jepun (Davies dan Stagg 1970).

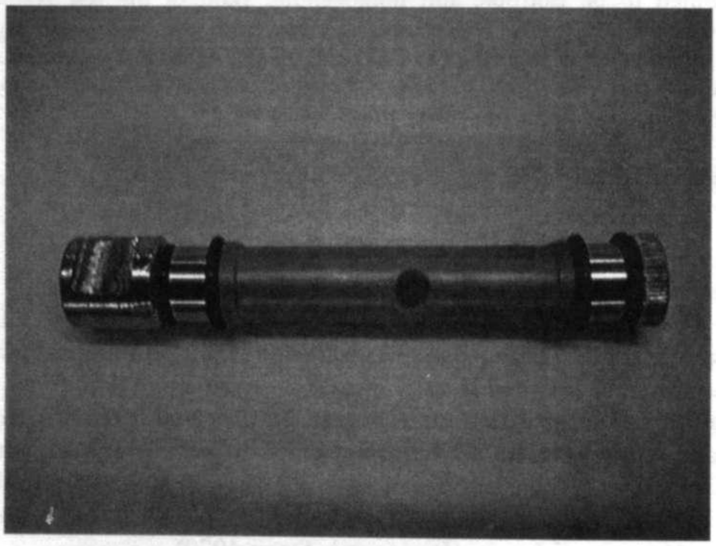
Ujian Brazilian dijalankan dengan memberi tekanan ke atas sampel ujian yang berbentuk cakera di antara dua plat pada mesin ujian yang sesuai sehingga sampel ujian pecah mengikut satah menegak. Tekanan pada kegagalan direkodkan dan dibahagikan dengan luas keratan permukaan akan memberikan kekuatan tegangan spesimen ujian. Saiz spesimen bagi ujian Brazilian adalah ketebalannya iaitu separuh daripada garis pusat.

Pengukuran kekuatan tegangan ke atas batuan lemah biasanya didapati susah untuk dilakukan disebabkan oleh keadaan asalnya adalah rapuh yang menyebabkan penyediaan sampel didapati susah untuk dilakukan. Alat H-Ometer dicipta sebagai satu alat ujian kekuatan tegangan tak langsung untuk batuan lemah sama ada di makmal ataupun di lapangan.

Tujuan kajian ini adalah untuk membandingkan keputusan ujian H-Ometer dengan ujian Brazilian. Rumusan perbandingan daripada keputusan kedua-dua ujian akan dibentuk.

H-OMETER

H-Ometer berbentuk selinder dengan lubang di bahagian dalam dan dipasang membran getah di sebelah luar yang boleh mengembang pada badannya *(Plat 1)*. Pada dasarnya ia mempunyai enam komponen yang utama iaitu Badan Utama, Kepala Sambungan, Klip Khas, Ekor, Tiub dan Membran. Ianya diperbuat daripada bahan keluli yang tahan karat. Kepala Sambungan adalah untuk menyambungkan Badan Utama dengan tiub disambungkan kepada unit kawalan. Unit kawalan ini adalah alat yang memberikan tekanan dengan air sebagai media utama yang merekodkan tekanan yang diberi dan isi padu air yang memasuki alat H-Ometer.



Plat 1: H-Ometer

Pam Digital Hidraulik digunakan sebagai unit kontrol untuk memberi tekanan dengan air suling (distilled water) sebagai media yang akan melalui tiub yang dipasang bagi menyambungkan pam dan H-Ometer. Tekanan yang diberi akan menolak air ke dalam H-Ometer dan ini akan membolehkan membran pada H-Ometer mengembang. Tekanan yang diberi akan direkod dan isi padu air yang memasuki H-Ometer akan disukat. Sebelum ujian dijalankan, proses penentukuran membran akan dijalankan terlebih dahulu dengan mengenakan tekanan ke atas H-Ometer di udara, yang merekodkan tekanan yang diberi dan isi padu air yang masuk. Penentukuran sistem juga dilakukan bagi memastikan tiada kebocoran berlaku. Setelah ujian dijalankan, keputusan yang didapati perlu diperbetulkan. Keputusan yang diperbetul adalah penting untuk mendapat nilai sebenar kekuatan bahan yang diuji.

PENYEDIAAN SAMPEL

Batuan lemah biasanya sukar untuk disediakan bagi tujuan pengujian di dalam makmal disebabkan oleh keadaan semula jadinya dan biasanya dikategorikan sebagai bahan yang bermasalah. Mengikut Meigh dan Wolski (1979) ada tiga kategori utama batuan lemah seperti ditunjukkan pada Jadual 1. Oleh itu bagi tujuan kajian ini, spesimen buatan telah digunakan. Untuk menghasilkan batuan buatan, kaedah campuran Bandis (1980) telah digunakan sebagai panduan. Pada tahun 1980, Bandis telah menghasilkan satu model bahan dengan ketetapan geometrik dan faktor skala tegasan yang mensimulasikan kandungan batuan yang sebenar. Penyediaan spesimen batuan buatan adalah berdasarkan bahan campuran Bandis dan dikeringkan pada suhu bilik. Untuk mendapatkan kekuatan yang sesuai bagi ujian H-Ometer, spesimen-spesimen ini telah diuji pada keadaan pengeringan yang berbeza.

JADUAL 1
Kategori batuan lemah (Meigh dan Wolski 1979)

Kategori	Ciri-Ciri
A. Baik	Kualiti spesimen baik, Kualiti lubang tebukan baik
B. Sederhana	Kualiti spesimen tidak baik, Kualiti lubang tebukan baik
C. Tidak baik	Kualiti spesimen tidak baik, Kualiti lubang tebukan tidak baik

Model batuan buatan terdiri daripada bahan-bahan berikut; pasir silika, barite, alumina kalsium, *plaster of Paris* dan air mengikut pecahan seperti yang ditunjukan pada Jadual 2. Barite digunakan sebagai cecair pelicin antara butiran-butiran pasir. Dalam kajian ini saiz butiran pasir yang digunakan adalah grit 30. Pasir, barite dan alumina kalsium dicampurkan secara kering dengan

JADUAL 2
Kandungan model batuan buatan mengikut pecahan

Bahan-Bahan	Pecahan
Pasir Silika	8
Barite	3
Alumina Kalsium	1
Plaster of Paris	1.26
Air	3

menggunakan mesin campuran mekanikal Hobart model A120 selama lima minit. Air dicampurkan dan penggaulan diteruskan selama lima minit lagi. Kemudian *plaster of Paris* dicampur dan digaulkan selama 2 minit. Akhirnya campuran dikeluarkan dan dituangkan ke dalam tiub selinder yang bersaiz 75mm garis pusat dan 150mm panjang. Ia dibiarkan selama empat minit sebelum dikeluarkan daripada tiub. Selepas empat minit, spesimen dikeluarkan dan dikeringkan pada suhu bilik.

Batuan buatan ini mempunyai warna keputihan seperti batu pasir dan mengikut Papalianggas (1986) ketumpatan spesimen ini adalah 1.85 g/cm3. Kekuatan spesimen ini bergantung kepada ikatan lemah oleh bahan asli yang bertindak sebagai simen antara butiran-butiran pasir tersebut.

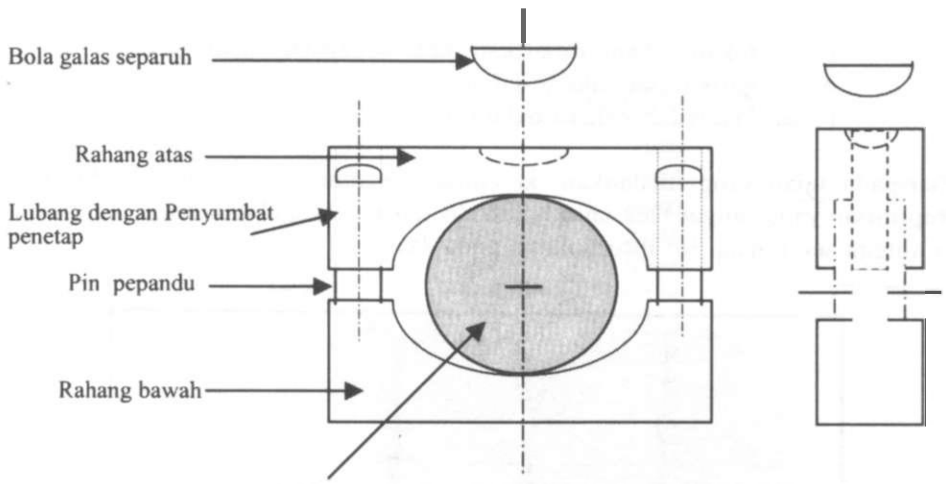
Untuk ujian H-Ometer, spesimen-spesimen ini ditebuk lubang dengan menggunakan mesin penebuk elektrik dengan mata penebuk bergaris pusat 10mm. Mata penebuk ini diperbuat daripada tungsten. Garis pusat lubang yang ditebuk pada spesimen berubah-ubah antara 10.2mm hingga 12mm. Ini disebabkan oleh ikatan pasir adalah ikatan rapuh yang mudah tanggal semasa proses penggerudian dilakukan. Semasa penggerudian dilakukan, keadaan yang berhati-hati perlu dititikberatkan. Dicadangkan kelajuan mesin gerudi adalah pada peringkat yang perlahan sesuai dengan keadaan spesimen.

Bagi ujian Brazilian, spesimen yang disediakan adalah sama dengan ujian H-Ometer. Selepas dikeringkan pada suhu bilik mengikut jangka masa sama dengan ujian H-Ometer, spesimen ini dipotong berbentuk selinder atau cakera dengan ketebalan separuh daripada garis pusat spesimen. Permukaan spesimen dilicinkan mengguna kepingan metal.

Sebanyak 30 spesimen telah disediakan iaitu 15 spesimen untuk setiap ujian. Tetapi didapati sejumlah 22 spesimen sahaja yang telah menepati piawaian bagi kedua-dua ujian. Sebanyak lapan spesimen lagi didapati rosak iaitu pecah sebelum ujian. Bagi tujuan kajian ini, keputusan ke atas 22 spesimen ini dibincangkan.

UJIAN BRAZILIAN

Ujian Brazilian merupakan kaedah tak langsung dalam menentukan kekuatan tegangan batuan di dalam makmal. Ujian pecahan ini dijalankan pada spesimen berbentuk selinder atau juga dipanggil cakera. Aparatus untuk ujian Brazilian ditunjukkan pada *Rajah 1* (ISRM 1981).



Rajah 1: Aparatus bagi ujian Brazilian (ISRM 1981)

Prosedur Ujian

Spesimen untuk ujian dipotong kepada bentuk selinder dengan ketebalannya adalah separuh daripada nilai ukuran garis pusat. Permukaan spesimen hendaklah bebas daripada sebarang tanda atau kesan seperti garisan semasa pemotongan dan jika ada ketidaksamaan pada permukaan, ia hendaklah tidak melebihi 0.025mm, dan kedua-dua belah permukaan hendaklah rata (ISRM 1981).

Pada ujian Brazilian, ketebalan (t) dan garis pusat (d) selinder spesimen diletakkan secara mendatar antara dua kepingan beban pada mesin ujian. Kemudian tekanan diberikan pada kedua-dua kepingan yang berlawanan arah sehingga spesimen ujian akan pecah dua mengikut satah diametrik tegak.

Pada kajian ini, spesimen yang disediakan adalah bergaris pusat 75mm dan ketebalannya adalah separuh daripada garis pusat iaitu 37.5mm. Spesimen berbentuk selinder ini disediakan dengan memotong batuan buatan menggunakan mesin gergaji bermata berlian dan air bersih digunakan sebagai pelincir semasa proses pemotongan. Permukaannya diratakan mengguna kepingan metal. Selepas mengambil bacaan ukuran ketebalan dan garis pusat spesimen, ujian dilakukan dengan menggunakan Mesin Ujian Universal 5 Ton. Mesin ini dilengkapi perekod carta untuk merekodkan bacaan tekanan pada kegagalan. Tekanan yang diberi adalah pada kadar yang tetap.

Keputusan Ujian Brazilian

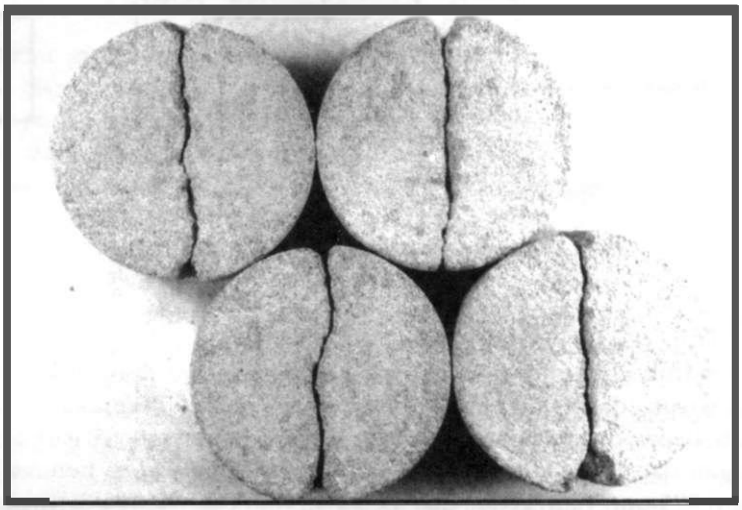
Kekuatan tegangan daripada ujian Brazilian diberikan oleh formula berikut:

$$\sigma_{t,B} = \frac{2P(1000)}{\pi\, d\, t}\,\mathrm{MPa} \qquad (1)$$

dengan,

P = tekanan yang diberikan pada kegagalan dalam kN
d = garis pusat cakera dalam mm
t = ketebalan cakera dalam mm

Daripada ujian yang dijalankan, kesemua spesimen yang diuji memberikan keputusan yang sangat baik iaitu boleh dikatakan kegagalan hampir sempurna sehingga sempurna. Ini ditunjukkan pada *Plat 2.*



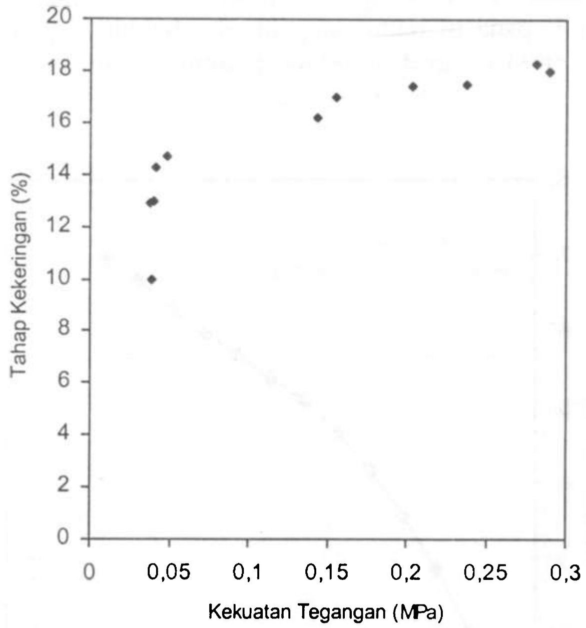
Plat 2: Kegagalan yang ditunjukkan oleh spesimen bagi ujian Brazilian

Keputusan ujian Brazilian ditunjukkan pada Jadual 3. Mengikut Bandis (1980), kekuatan tegangan daripada ujian Brazilian ke atas batuan buatan menggunakan model campuran Bandis adalah di antara 0.127 ke 0.473 MPa bergantung kepada kadar pengeringan spesimen. Daripada Jadual 3, kekuatan tegangan batuan buatan yang dikeringkan pada suhu bilik adalah antara 0.038 hingga 0.289 MPa. Nilai ini didapati rendah berbanding nilai yang diberikan oleh Bandis (1980). Ini adalah disebabkan spesimen dikeringkan pada suhu bilik dan diuji pada tahap kekeringan yang rendah berbanding Bandis (1980) yang ujian dilakukan setelah spesimen dikeringkan pada suhu bilik selama 24 jam dan dimasukkan ke dalam ketuhar pada suhu 85°C selama 36 jam lagi.

Keputusan kekuatan tegangan ini dibandingkan dengan nilai tahap kekeringan spesimen dan ditunjukkan pada *Rajah 2.* Hasil keputusan mendapati peratusan kekeringan bertambah, nilai kekuatan tegangan juga bertambah. Ini kerana spesimen lebih kering mempunyai ikatan antara butiran lebih kuat dan memberikan nilai kekuatan tegangan lebih tinggi.

JADUAL 3
Kekuatan tegangan daripada ujian Brazilian

Tahap kekeringan (%)	Kekuatan Tegangan (MPa)
10.0	0.039
12.9	0.038
13.0	0.040
14.3	0.041
14.7	0.048
16.2	0.144
17.0	0.155
17.4	0.203
17.5	0.238
18.0	0.289
18.3	0.281



Rajah 2: Peratusan kekeringan melawan kekuatan tegangan dan ujian Brazilian

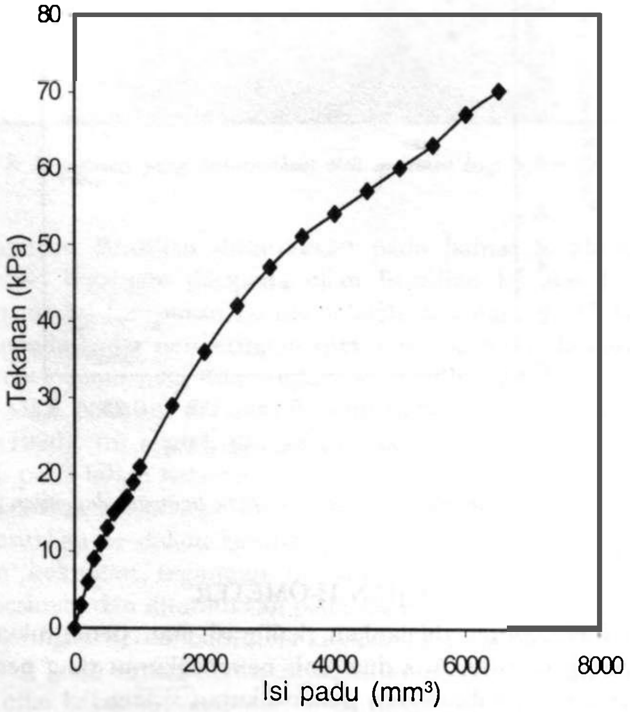
UJIAN H-OMETER

Sebelum ujian H-Ometer dijalankan, kalibrasi atau penentukuran perlulah dilakukan terlebih dahulu. Ada dua jenis penentukuran yang perlu dilakukan iaitu penentukuran membran dan penentukuran sistem.

Penentukuran

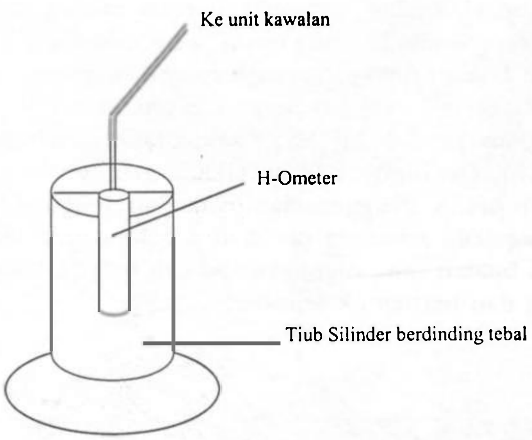
Penentukuran adalah sangat penting dilakukan pada ujian H-Ometer. Ini adalah kerana graf lengkuk tekanan-isi padu daripada ujian perlu diperbetulkan supaya keputusan kekuatan tegangan yang didapati daripada ujian H-Ometer adalah kekuatan yang sebenar. Mair dan Wood (1987) menerangkan walaupun kecil nilai perubahan daripada pembetulan yang dilakukan ke atas keputusan ujian akan memberikan kesan yang besar ke atas nilai moduli yang didapati daripada ujian yang dijalankan pada tanah keras dan batuan lemah.

Oleh itu untuk H-Ometer, penentukuran membran dilakukan dengan mengembangkan membran di udara dengan memberikan isi padu air yang memasuki H-Ometer secara tetap (katakan $100mm^3$) dan tekanan direkodkan. Ini penting dilakukan dalam proses membetulkan nilai bacaan yang didapati daripada ujian dan ia juga dipanggil sebagai pembetulan membran. Semasa penentukuran dijalankan, H-Ometer diletakkan secara menegak supaya sama seperti ianya diletakkan ke dalam sampel semasa ujian dijalankan. Bagi bacaan yang diambil, Clarke (1995) mengesyorkan supaya ianya dibaca setelah dibiarkan selama satu minit pada isi padu yang diberi sebelum isi padu ditambahkan. *Rajah 3* menunjukkan graf lengkuk penentukuran membran daripada H-Ometer.

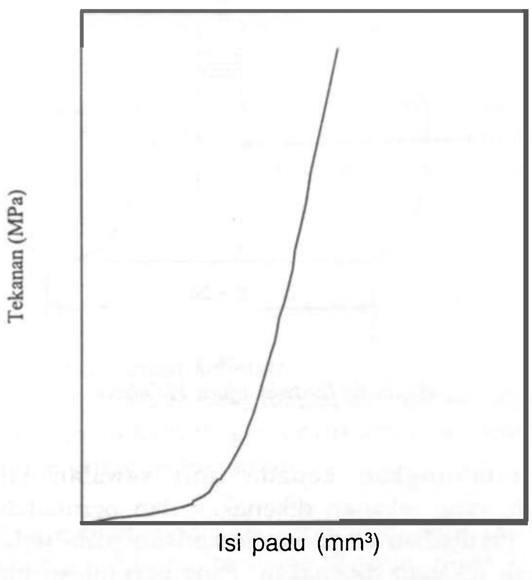


Rajah 3: Graf lengkuk penentukuran membran daripada H-Ometer

Penentukuran sistem dilakukan dengan memasukkan H-Ometer ke dalam tiub selinder berdinding tebal seperti pada *Rajah 4.* Tekanan diberikan kepada H-Ometer, yang akan mengembangkan membran lalu menolak kepada dinding tebal tiub selinder sehingga mencapai tahap tekanan maksimum. Setiap tekanan yang diberi akan dibiarkan selama satu minit sebelum bacaan direkodkan. Penentukuran ini penting dilakukan bagi memastikan sistem H-Ometer bebas daripada kebocoran. *Rajah 5* menunjukkan graf lengkuk penentukuran sistem.



Rajah 4: H-Ometer dimasukkan ke dalam tiub silinder berdinding tebal sebagai penentukuran sistem



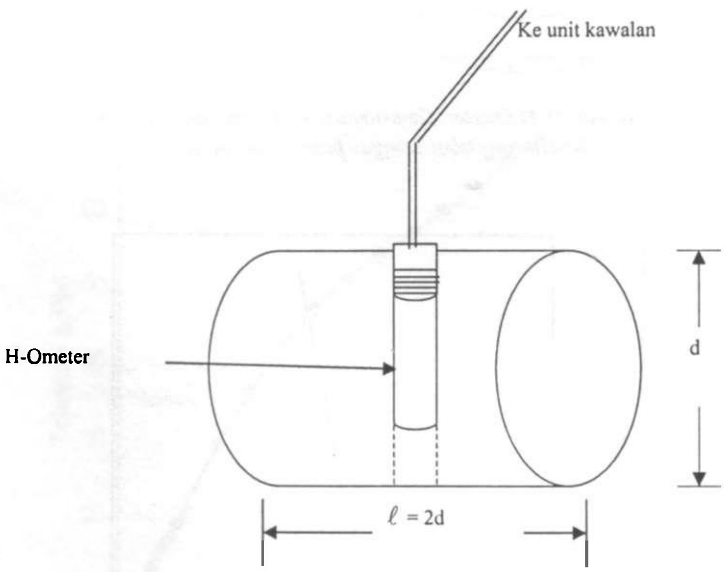
Rajah 5: Graf lengkuk penentukuran sistem

Sebagai peraturan am, ujian penentukuran perlu dilakukan setiap kali sebelum ujian dijalankan.

Prosedur Ujian

H-Ometer merupakan ujian tegasan terkawal kerana kenaikan tekanan yang dikenakan adalah dengan kadar yang telah ditetapkan. Kadar kenaikan tekanan yang dikenakan dianggarkan sehingga kegagalan pada spesimen berlaku ataupun sehingga had tekanan, bergantung kepada kekuatan bahan yang diuji.

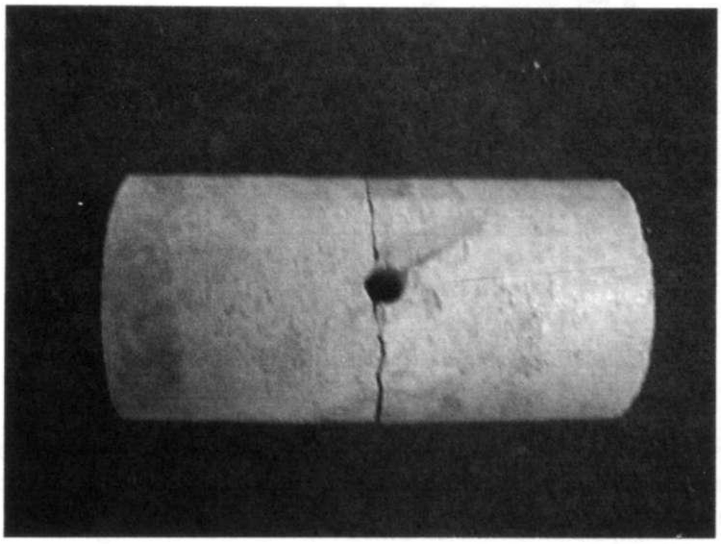
Pada prinsipnya, H-Ometer dimasukkan pada lubang yang telah ditebuk pada spesimen ujian (batuan buatan) untuk ujian dijalankan *(Rajah 6)*. Lubang ujian pada batuan buatan ditebuk menggunakan mesin gerudi elektrik secara tegak dengan mata gerudi tungsten bergaris pusat 10mm digunakan. Biasanya susah untuk mendapatkan lubang yang ditebuk tadi betul-betul sesuai ataupun padat dengan H-Ometer. Ini disebabkan oleh batuan lemah merupakan bahan rapuh yang mudah pecah. Penggerudian dilakukan dengan berhati-hati supaya sampel tidak mengalami sebarang pecahan sebelum ujian dijalankan. Dalam kajian ini, batuan buatan yang digunakan adalah bersaiz 150mm panjang dan 75mm garis pusat dan berbentuk selinder.



Rajah 6: Ilustrasi ujian H-Ometer

H-Ometer disambungkan kepada unit kawalan iaitu pam digital menggunakan tiub, yang tekanan dikenakan dan perubahan isi padu dalam mm^3 direkodkan. Perubahan isi padu direkodkan pada setiap 30 saat dan 60 saat bagi setiap kali tekanan dikenakan. Pada permulaan ujian, pertambahan tekanan pada kadar 25 kPa ditetapkan sehingga tekanan mencapai 100 kPa.

Selepas 100 kPa, pertambahan tekanan dinaikkan antara 50 hingga 100 kPa bergantung kepada kesesuaian sehingga spesimen gagal sepenuhnya. Kegagalan berlaku dengan tiba-tiba dan tanpa amaran. Oleh itu, sangatlah penting untuk memerhatikan kadar perubahan isi padu semasa ujian dijalankan dengan teliti. Pertambahan tekanan sebanyak 50 hingga 100 kPa dicadangkan bagi menjaga keselamatan semasa ujian dan juga bagi mendapatkan keputusan yang baik. *Plat 3* menunjukkan spesimen gagal selepas ujian H-Ometer.



Plat 3: Spesimen (batuan buatan) pecah kepada dua selepas ujian H-Ometer

Keputusan Ujian H-Ometer

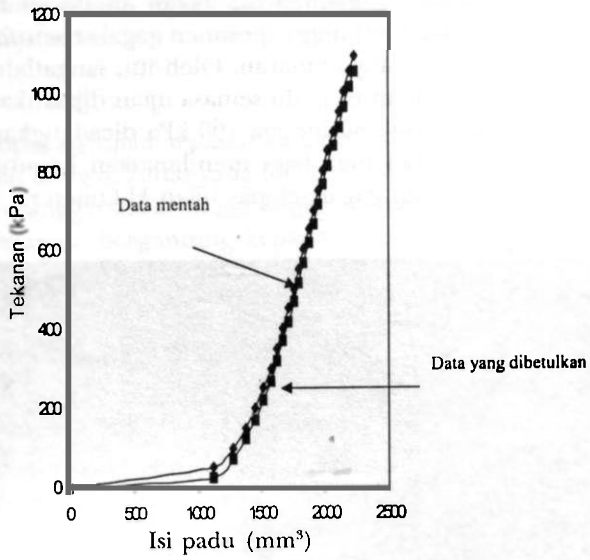
Keputusan yang didapati daripada ujian H-Ometer tidak berguna jika tidak diperbetulkan. Oleh itu, keputusan ujian atau data mentah perlu dibetulkan bagi mendapatkan tekanan sebenar. Tekanan sebenar dikira menggunakan formula berikut:

$$P_{corr} = P_a - P_{cal} \tag{2}$$

iaitu,

P_{corr} adalah tekanan sebenar
P_a adalah tekanan yang dikenakan semasa ujian
P_{cal} adalah tekanan penentukuran daripada graf lengkuk penentukuran membran

Tekanan sebenar yang didapati setelah pembetulan dibuat seterusnya digunakan untuk menghasilkan graf tekanan melawan isi padu *(Rajah 7)*. Daripada graf ini, tekanan sebenar pada kegagalan, P_f, dan tekanan awal semasa membran



Rajah 7: Graf lengkuk tekanan-isi padu dari ujian H-Ometer

menyentuh dinding pada spesimen ujian, P_o, akan diperolehi dan ini akan digunakan untuk mengira kekuatan tegangan. Kekuatan tegangan boleh dikira menggunakan formula berikut:

$$\sigma_{t,\,HO} = P_f - 4\,P_o \qquad (3)$$

iaitu,

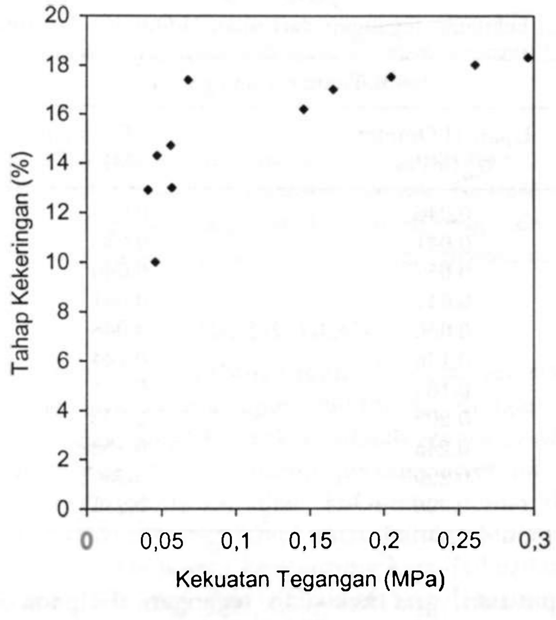
$\sigma_{t,\,HO}$ adalah kekuatan tegangan dari ujian H-Ometer
P_f adalah tekanan pada kegagalan
P_o adalah tekanan awal

Keputusan kekuatan tegangan spesimen yang diuji ditunjukkan pada Jadual 4. Tekanan pada kegagalan, P_f, dan tekanan awal, P_o, adalah dua nilai yang sangat penting untuk mendapatkan nilai kekuatan tegangan menggunakan ujian H-Ometer. Tekanan awal ditentukan dari graf keputusan ujian. Daripada keputusan kekuatan tegangan yang diperoleh, satu kaitan di antara tahap kekeringan spesimen dan nilai kekuatan tegangan dibentuk. Ini ditunjukkan pada *Rajah 8.* Hasil keputusan menunjukkan peratusan kekeringan bertambah, nilai kekuatan tegangan juga bertambah.

Keputusan nilai kekuatan tegangan dari ujian H-Ometer didapati sedikit besar berbanding nilai kekuatan tegangan yang diberikan oleh Bandis (1980). Ini adalah kerana tahap pengeringan yang berbeza berbanding spesimen yang disediakan oleh Bandis (1980).

JADUAL 4
Kekuatan tegangan daripada ujian H-Ometer

Tahap Kekeringan (%)	Kekuatan Tegangan (MPa)
10.0	0.046
12.9	0.041
13.0	0.057
14.3	0.047
14.7	0.056
16.2	0.146
17.0	0.166
17.4	0.208
17.5	0.245
18.0	0.299
18.3	0.296



Rajah 8: Peratusan kekeringan melawan kekuatan tegangan dari ujian H-Ometer

PERBANDINGAN KEKUATAN TEGANGAN DARIPADA UJIAN H-OMETER DAN UJIAN BRAZILIAN

Adalah menjadi kebiasaan untuk membandingkan keputusan yang didapati dari ujian alatan baru dengan keputusan dari alat ujian yang sedia ada, sama ada ujian dijalankan di lapangan ataupun di makmal. Daripada perbandingan satu kaitan baru biasanya dapat dibentuk dan perbandingan juga dapat memastikan sama ada alatan baru sesuai digunakan ataupun tidak.

Daripada kajian ini, perbandingan dilakukan dengan ujian Brazilian. Mengikut Clarke (1995), tiada sebab mengapa keputusan dari alat meter tekanan semestinya sama dengan ujian lain yang dilakukan. Banyak perbandingan daripada ujian-ujian yang telah dijalankan dalam bidang geoteknik seperti diberikan oleh Briaud (1986) dan Wilson dan Corke (1990). Biasanya didapati sangat sedikit keputusan perbandingan memberikan keputusan yang sama tetapi perbandingan boleh menilai sesuatu keputusan itu sama ada berkualiti ataupun tidak.

Keputusan kedua-dua ujian yang dijalankan diringkaskan dan ditunjukkan pada Jadual 5 di bawah. Kedua-dua ujian telah dilakukan ke atas spesimen yang dikeringkan pada tahap yang dikehendaki dan diuji pada masa yang sama. Keputusan daripada ujian H-Ometer memberikan nilai yang lebih tinggi berbanding nilai daripada ujian Brazilian. Walaupun ada sedikit perbezaan, keputusan adalah sangat baik.

JADUAL 5
Perbandingan kekuatan tegangan dari ujian H-Ometer dan ujian Brazilian

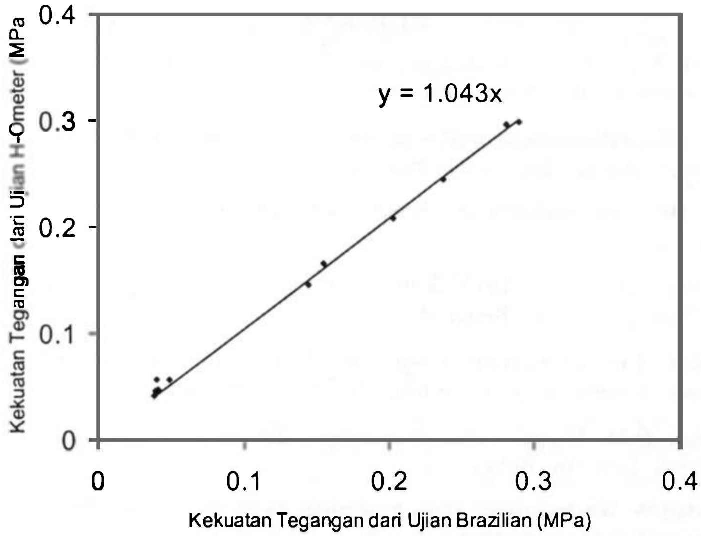
Kekuatan Tegangan	
Ujian H-Ometer σ_{ho}(MPa)	Ujian Brazilian σ_{b}(MPa)
0.046	0.039
0.041	0.038
0.057	0.040
0.047	0.041
0.056	0.048
0.146	0.144
0.166	0.155
0.208	0.203
0.245	0.238
0.299	0.289
0.296	0.281

Daripada keputusan, graf kekuatan tegangan daripada ujian H-Ometer melawan kekuatan tegangan daripada ujian Brazilian telah dibentuk dan ini ditunjukkan pada *Rajah 9.* Graf menunjukkan satu kaitan lelurus yang jelas di antara nilai daripada ujian H-Ometer dan ujian Brazilian. Hasil perbandingan ini, satu kaitan dicadangkan adalah seperti berikut:

$$\sigma_{t,HO} = 1.043\ \sigma_{t,B} \qquad (4)$$

dengan,

$\sigma_{t,HO}$ adalah kekuatan tegangan daripada ujian H-Ometer
$\sigma_{t,B}$ adalah kekuatan tegangan daripada ujian Brazilian



Rajah 9: Kekuatan tegangan dari ujian H-Ometer melawan kekuatan tegangan dari ujian Brazilian

Daripada keputusan di atas, jelaslah bahawa ujian H-Ometer dapat digunakan untuk mendapatkan nilai kekuatan tegangan ke atas batuan lemah. Walaupun perbandingan di atas adalah secara ghalib, keputusan yang didapati menunjukkan ujian H-Ometer adalah berkait rapat dengan ujian Brazilian.

KESIMPULAN

Dari hasil keputusan dan perbincangan di atas, jelaslah bahawa satu perbandingan yang baik antara ujian H-Ometer dengan ujian Brazilian. Walaupun perbandingan yang dijalankan adalah secara ghalib antara kedua-dua ujian, keputusan yang didapati menunjukkan persetujuan yang amat baik antara kedua-dua ujian. Oleh itu ujian H-Ometer sesuai digunakan untuk mendapatkan nilai kekuatan tegangan pada batuan lemah. Nilai kekuatan tegangan daripada ujian H-Ometer bergantung kepada keadaan spesimen atau batuan lemah yang diuji. Pada kajian ini, batuan lemah buatan telah digunakan kerana ia memberikan spesimen yang diuji berada dalam keadaan yang isotropik dan homogen, jadi ia tidak mengandungi apa-apa ketakselanjaran dan ini menunjukkan spesimen mempunyai ciri-ciri yang elastik. Hal ini penting untuk mendapatkan keputusan yang seragam yang akhirnya dapat digunakan untuk proses perbandingan antara kedua-dua ujian.

PENGHARGAAN

Penulis ingin merakamkan ucapan terima kasih kepada Kementerian Sains, Teknologi dan Alam Sekitar kerana membiayai projek penyelidikan ini di bawah geran IRPA (vot: 03-02-04-0083).

RUJUKAN

BANDIS, S. 1980. Experimental studies of scale effects on shear strength and deformation of rock joints. Ph.D. Thesis, University of Leeds, U.K.

BRIAUD, J.L. 1986. Pressuremeter and deep foundation design. In *Proc. 2nd. Int'l Symp. on Pressuremeter Marine Appl.*, p. 376-405. ASTM STP 950, Texam, USA.

CLARKE, B.G. 1995. *Pressuremeters in Geotechnical Design.* Glasgow: Blackie Academic and Professional.

DAVIES, J.D. dan K.G. STAGG. 1970. Splitting tests on rock specimens. In *Proc. of 2nd. Congr. ISRM*, p. 343-349. Beograd.

ISRM. 1981. Rock Characterization, Testing and Monitoring. *ISRM Suggested Methods* ed. E.T. Brown. Commission on Testing Methods. Pergamon.

MAIR, R.J. dan D.M. WOOD. 1987. *Pressuremeter Testing – Methods and Interpretation.* Butterworth, London: Butterworth.

MEIGH, A.C. dan W. WOLSKI. 1979. Design parameters for weak rocks. In *Proc. of 7th. Eur. Conf. SMFE*, Brighton. **5:** 59-79.

PAPALIANGAS, T. 1980. The effect of frictional filling materials on shear behaviour of rock discontinuities. M.Sc. Thesis, University of Leeds, U.K.

WILSON, W. dan D.J. CORKE. 1990. A comparison of modulus values of sandstone derived from high pressure dilatometer, plate bearing, geophysical and laboratory testing. In *Proc. of 3rd. Int'l Symp. on Pressuremeter*, p. 351-360. Oxford.

Pertanika J. Sci. & Technol. **10(1)**: 89 - 97 (2002)

ISSN: 0128-7680

Model Resapan Mekanik Kuantum Zarah Bebas dalam Selang Terbatas

Shaharir bin Mohamad Zain & Nik Rusdi bin Yaacob
Pusat Pengajian Sains Matematik
Universiti Kebangsaan Malaysia
43600 UKM, Bangi
Selangor, Malaysia

Diterima: 8 Januari 2001

ABSTRAK

Mengikut tafsiran piawai (tafsiran Born dan tafsiran Copenhagen Bohr) mekanik kuantum, magnitud kuasa dua fungsi gelombang bagi sesuatu zarah selama ini dianggap mewakili fungsi ketumpatan kebarangkalian bagi zarah berkenaan berada dalam keadaan tersebut apabila sukatan dilakukan ke atasnya. Dalam makalah ini, ditunjukkan bahawa bagi zarah yang secara klasiknya berupa zarah bebas (potensinya sifar) berjisim m dalam selang terbatas, kuasa dua modulus fungsi gelombangnya memenuhi persamaan resapan dengan pekali resapan $\rho = \hbar/2m$, pekali hanyutan $-2\frac{\partial F}{\partial x}$ dan bersumberkan $W = 2(U + \frac{1}{4}\kappa^2)$ dengan F ialah fasa fungsi gelombang berkenaan, κ suatu ungkapan yang melibatkan fasa dan tenaga zarah tersebut, dan $U = \rho\frac{\partial^2 F}{\partial x^2} + \frac{\partial F}{\partial t} + \rho\left(\frac{\partial F}{\partial x}\right)^2$ ialah sumber bagi resapan yang berpekali resapan dan hanyutan yang sama.

ABSTRACT

Acoording to the standard interpretation of quantum mechanics (Born and Bohr-Copenhagen interpretations), the square of the magnitude of the wave function represents a probability density function for a particle in a state whereby a measurement is made on it. In this article it is shown that a quantum particle which is classically a free particle in a bounded interval with mass *m*, has its wave function whose square modulus satisfies a diffusion equation with a diffusion coefficient $\rho = \hbar/2m$, a drift coefficient $-2\frac{\partial F}{\partial x}$ and a source $W = 2(U + \frac{1}{4}\kappa^2)$ where F is the phase function, κ depends on the phase and the energy of the particle, and $U = \rho\frac{\partial^2 F}{\partial x^2} + \frac{\partial F}{\partial t} + \rho\left(\frac{\partial F}{\partial x}\right)^2$ is a source of a diffusion with the same diffusion and drift coefficients.

PENGENALAN

Semenjak kemunculan persamaan Schroedinger (1926) yang terkenal itu, selepas ini dirujuk sebagai PsS sahaja, iaitu

$$\hbar k = \frac{\partial \Psi}{\partial t} = (-\frac{\hbar^2}{2m}\nabla^2 + V)\Psi, k = \sqrt{-1}), \ \hbar \text{ pemalar Planck-Dirac,} \tag{1}$$

bagi zarah klasik berjisim m dan berpotensi V, pelbagai tafsiran telah diberikan kepada fungsi gelombang Ψ sehingga melahirkan berbagai-bagai mazhab dalam mekanik kuantum. Tafsiran yang paling terkenal ialah tafsiran Born (1926), iaitu $|\Psi|^2$ ialah fungsi ketumpatan kebarangkalian bagi zarah berkenaan itu berada dalam sesuatu kedudukannya apabila sukatan dilakukan ke atasnya. Tafsiran ini cuba diperjelaskan oleh ramai pihak dan sekaligus mendapat tentangan daripada pelbagai pihak yang lain sehingga melahirkan banyak mazhab dalam mekanik kuantum (Lihat keadaan ini umpamanya dalam Wick 1995). Hujah-hujah Born yang asal itu telah cuba diperjelaskan oleh ramai penyokongnya dari semasa ke semasa hingga hari ini seperti yang dapat dilihat dalam Pais (1982), Cartwright (1987) dan Robinett (1997) kerana tafsiran ini (yang terkenal sebagai tafsiran Copenhagen) masih menjadi ikutan umum walaupun kelemahannya juga diutarakan dari semasa ke semasa (sejak Einstein mengkritik keras Bohr dan rakan-rakan sealirannya di Copenhagen itu) dan kritikan yang panjang lebar dan canggih pada dua dasawarsa terakhir ini ialah yang dilakukan oleh Healey (1990). Keserupaan PsS dengan persamaan resapan, selepas ini yang kemudian itu dirujuk sebagai PsR sahaja,

$$\frac{\partial w}{\partial t} = (r\nabla^2 + U)w \tag{2}$$

dengan $\rho > 0$ sebagai pekali resapan yang boleh bergantung pada masa dan U sebagai sumber resapan, telah mendorong ramai pihak mengkaji PsS menerusi PsR kerana, antara lainnya, PsR berhubung rapat dengan proses stokastik yang taburannya diberikan oleh penyelesaian PsR itu, dan membawa kepada penyelesaian Kac dan Feynman yang sorotan mutakhirnya dapat dilihat dalam tesis Zainal (1997). Sementara itu Bess (1994) cuba berhujah menerusi mekanik statistik klasik dan persamaan Jakobi-Hamilton dalam mekanik zarah klasik bagi memperoleh PsS yang bertujuan menzahirkan asal-usul kebarangkalian dalam mekanik kuantum itu. Karya yang lebih meyakinkan dan lebih komprehensif dan luas jangkauannya ialah kerja Nagasawa (1993) dan Aebi (1996), tetapi mereka ini memulakan kajian analogi PsS dan konjugatnya dengan sepasang PsR yang satu daripadanya berpekali resapan negatif, yang tentunya menimbulkan banyak kemusykilan. Yang lainnya berpuas hati dengan model mekanik bendalir klasik dan hidrodinamik klasik dengan mengimbau kepada persamaan keselanjaran dan persamaan Euler dalam bidang ini, iaitu suatu kajian yang dimulai oleh Madelung (1926) dan mendapat sambutan yang hebat juga sehingga kini seperti yang dapat dilihat dalam karya Gilson (1978, 1980), Sonego (1991) dan Wallstrom (1994). Satu kumpulan lagi, yang mendapat inspirasi dan galakan daripada Einstein, ialah yang langsung ingin menyingkir tafsiran kebarangkalian dalam mekanik kuantum kerana mereka tidak percaya kepada hukum alam (tabii atau *sunnatullah*) yang secara intrinsiknya berkebarangkalian (atau kini didapati lebih am lagi hukum kemungkinan menerusi model kabur itu). Sarjana yang paling berjaya dalam pendekatan ini ialah Bohm

(1952, 1984) yang kini melahirkan mekanik yang dinamai mekanik Bohman (Berndl *et al.* 1995, Cushing *et al.* 1996) dan implikasi falsafah dan metafiziknya cukup besar seperti yang dapat dilihat dalam Bohm *et al.* (1987), Bohm dan Hiley (1993) dan Sharpe (1993).

Kajian kami ini juga dimotivasikan oleh analogi PsS dengan PsR tetapi berbeza daripada karya-karya yang tersebut di atas, kerana kami menumpukan kepada sifat $|\Psi|^2$ modulus kuasa dua fungsi gelombang itu, dan buat permulaan ini kepada kasus zarah bebas (iaitu potensi V = 0) dalam selang terhingga (a,b) (sehingga tenaganya menjadi diskret dan positif). Kami membuktikan kuantiti ini memenuhi persamaan resapan dan resapan kuasa dua iaitu persaman untuk bagi u memenuhi PsR, persamaan (2) itu. Kajian ini ialah pelengkap kepada kajian kami sebelum ini (Shaharir dan Nik Rusdi 2000) yang menyelesaikan isu ini menerusi penyelesaian kamiran (daripada teknik jelmaan Fourier dan bukan pemisahan pemboleh ubah yang dilakukan di sini), walaupun tersirat dalam penyelesaian kamiran (menerusi jelmaan Fourier itu) terkandung kasus tenaga diskret ini. Kami sedar tentang betapa istilah zarah bebas yang biasanya dimaksudkan sebagai zarah berpotensi sifar atas seluruh set nombor nyata yang mengimplikasikan mekanik kuantumnya menghasilkan hanya tenaga positif yang kontinum dan oleh itu sudah pun dibicarakan dalam makalah kami (2000) sebelum ini.

Teorem 1

Apabila Ψ memenuhi PsS zarah bebas berjisim m dengan syarat awalnya $\Psi(\bullet,0) = \phi(x)$, fungsi yang kuasa duanya terkamirkan Lebesgue, dan zarah atom itu berada dalam selang terbatas (a,b) sehingga $\Psi(a,\bullet) = \Psi(b,\bullet)$, maka kuasa dua modulus fungsi gelombang $S = |\Psi|^2$ memenuhi persamaan resapan yang berpekali resapan $\rho = \frac{\hbar}{2m}$ dan berpekali hanyutan $-2\rho\frac{\partial F}{\partial x}$ serta bersumberkan $W=2(U + \frac{1}{4}\rho\kappa^2)$ yang

$$F = \tan^{-1}\left[\frac{\sum|c_r|\sin[(p_r x - E_r t + \varphi_r)/\hbar]}{\sum|c_r|\operatorname{kos}[(p_r x - E_r t + \varphi_r)/\hbar]}\right] \text{ fasa fungsi gelombang,}$$

$$\kappa = \frac{2\sum|c_r c_s|p_r \sin(f_{rs})}{\hbar\sum|c_r c_s|\operatorname{kos}(f_{rs})},\ f_{rs} = [(p_r - p_s)x + E_r - E_s)t + \varphi_r - \varphi_s]/\hbar,\ \varphi_r = \hbar\ \mathrm{huj}(c_r),$$

c_n ialah pekali Fourier kompleks bagi fungsi ϕ gelombang awal, E_r ialah aras tenaga ke-r, $p_r^2 = 2mE_r$ dan $U = \rho\frac{\partial^2 F}{\partial x^2} + \frac{\partial F}{\partial t} + \rho\left(\frac{\partial F}{\partial x}\right)^2$ ialah sumber bagi resapan yang dipenuhi oleh $R = |\Psi|$ yang berpekali resapan dan hanyutan yang sama dengan resapan S itu.

Bukti:

Penyelesaian PsS zarah bebas dan dalam selang terbatas (a,b) ialah

$$\Psi(x,t) = \sum_{n=-\infty}^{\infty} |c_n| \mathrm{eksp}[(<\mu_n, q> + \varphi_n)k/\hbar] \tag{i}$$

$$\varphi_n / \hbar = \mathrm{huj}(c_n) \tag{ii.a}$$

$$\mu_n = (p_n, E_n), q = (x,t) \tag{ii.b}$$

$$p_n^2 = 2mE_n = \left(\frac{n\pi\hbar}{b-a}\right), E_n > O,\ n,\ \text{integer} \tag{ii.c}$$

dan

$$<\mu, q) = px - Et, \tag{ii.d}$$

$$c_n = \frac{1}{\|\Psi_n\|^2} \int_a^b \varphi(x) \mathrm{eksp}[-kp_n x/\hbar] dx \tag{ii.e}$$

yang

$$\Psi_n(x) = \mathrm{eksp}[kp_n x / \hbar] \tag{ii.f}$$

$$\|f\|^2 = \int_a^b |f(x)|^2 dx \tag{ii.g}$$

dan

$$\int_a^b \Psi_n(x)\overline{\Psi}_m(x)dx = \|\Psi_n\|^2 \delta_{mn} \tag{ii.h}$$

[Dalam buku-buku teks, selang yang dipertimbangkan ialah selang simetri (-b,b). Lihat umpamanya Schiff 1968]

Dalam bentuk perwakilan kutub, fungsi gelombang ini ialah

$$\Psi = Re^{kF} \tag{iii}$$

dengan

$$R^2 = S = \sum |c_n c_r| \text{eksp}(kf_{nr}) \tag{iv.a}$$

$$f_{nr} = L_n - L_r \tag{iv.b}$$

$$L_n = [\langle \mu_n, q \rangle + \varphi_n]/\hbar = (p_n x - E_n t + \varphi_n)/\hbar, \mu = (x, t) \tag{iv.c}$$

$$\varphi_n = \text{huj}(c_n) \tag{iv.d}$$

dan

$$F = \tan^{-1}\left[\frac{\sum |c_n| \sin(L_n)}{\sum |c_n| \text{kos}(L_n)}\right] \tag{v}$$

Oleh itu

$$\frac{\partial S}{\partial x} = \frac{1}{\hbar} k \sum |c_n c_r| \text{eksp}[kf_{nr}](E_r - E_n) \tag{vi.a}$$

$$= -\frac{2}{\hbar} \sum |c_n c_r| \sin(f_{nr}) E_r, \text{ sebutan kos itu menjadi sifar,} \tag{vi.b}$$

dan

$$\frac{\partial S}{\partial x} = \frac{1}{\hbar} k \sum |c_n c_r| \text{eksp}[kf_{nr}](p_n - p_r) \tag{vii.a}$$

$$= -\frac{2}{\hbar} \sum |c_n c_r| \sin(f_{nr}) p_n \text{ , sebutan kos menjadi sifar.} \tag{vii.b}$$

Oleh itu

$$\frac{\partial^2 S}{\partial x^2} = -\frac{1}{\hbar^2} \sum |c_n cr| \text{eksp}[f_{nr}](p_n - p_r)^2 \tag{viii.a}$$

$$= -\frac{1}{\hbar^2} \sum |c_n c_r| \text{eksp}[kf_{nr}](2m(E_n + E_r) - 2p_n p_r) \tag{viii.b}$$

$$= -\frac{2}{\hbar^2} \sum |c_n c_r| \text{kos}(f_{nr})(2mE_n - p_n - p_r) \tag{viii.c}$$

Dengan ini, maka

$$\frac{\partial S}{\partial t} - \gamma \frac{\partial^2 S}{\partial x^2} = -\frac{2}{\hbar} \Sigma |c_n c_r| \sin(f_{nr}) E_r + \frac{\gamma}{\hbar^2} \Sigma |c_n c_r| \mathrm{kos}(f_{nr}) (4mE_r - 2p_n p_r) \qquad \text{(ix)}$$

Jelaslah $\gamma = \frac{\hbar}{2m}$ ialah nilai pekali resapan yang paling bersahaja lagi memenuhi keperluan unit atau matra jasmani bagi persamaan di atas.

Tetapi

$$\frac{\partial F}{\partial x} = \frac{1}{\hbar S} \Sigma |c_n c_r| p_n \mathrm{kos}(f_{nr}) \qquad \text{(x)}$$

Oleh itu

$$\frac{\partial^2 F}{\partial x^2} = -\frac{1}{s} \frac{\partial S}{\partial x} \frac{\partial F}{\alpha x} - \frac{1}{\hbar^2 S} \Sigma |c_n c_r| p_n (p_r - pn) \sin(f_{rn})$$

atau

$$\frac{\partial F}{\partial x} \frac{\partial S}{\partial x} + \frac{\partial^2 F}{\partial x^2} S = \Sigma |c_n c_r| \left(\frac{2mE_n}{\hbar^2} \right) \sin(f_{rn}) \qquad \text{(xi)}$$

setelah menggunakan identiti

$$\Sigma |c_n c_r| p_n p_r \sin(f_{rn}) = 0 \qquad \text{(xii)}$$

kerana $f_{rn} = -f_{nr}$ sedangkan faktor lainnya simetri.

Dengan hasil (xi) ini, maka persamaan (ix) menjadi

$$\frac{\partial S}{\partial t} - \gamma \frac{\partial^2 S}{\partial x^2} = -2\rho \frac{\partial F}{\partial x} \frac{\partial S}{\partial x} - 2\rho \frac{\partial^2 F}{\partial x^2} S + \frac{\gamma}{\hbar^2} \Sigma |c_n c_r| \mathrm{kos}(f_{nr}) (4mE_r - 2p_n p_r) \qquad \text{(xiii)}$$

Tetapi

$$\frac{\partial S}{\partial t} = -\frac{1}{\hbar S} \Sigma |c_n c_r| E_r \mathrm{kos}(f_{nr}) \qquad \text{(xiv)}$$

Oleh itu persamaan (xiii) menjadi

$$\frac{\partial S}{\partial t} - \gamma\frac{\partial^2 S}{\partial x^2} = -2\left[\rho\frac{\partial F}{\partial x}\frac{\partial S}{\partial x} + \rho\frac{\partial^2 F}{\partial x^2}S + \frac{\partial F}{\partial t}S\right] - \frac{2\gamma}{\hbar^2}\sum\left|c_n c_r\right|kos(f_{nr})p_n p_r \qquad (xv)$$

Tetapi daripada persamaan (x)

$$\left(\frac{\partial F}{\partial x}\right)^2 = \frac{1}{\hbar^2 S^2}\sum\sum\left|c_n c_r\right|\left|c_u c_v\right|p_n p_u kos(f_{nr})kos(f_{uv})$$

$$= \frac{1}{\hbar^2 S^2}\sum\sum\left|c_n c_r c_u c_v\right|p_n p_u \{kos(f_{nu})\ kos(f_{rv}) - sin(f_{nr})sin(f_{uv})\}$$

$$= \frac{1}{\hbar^2 S}\sum\left|c_n c_u\right|p_n p_u kos(f_{nu}) - \frac{1}{\hbar S}\sum\left|c_n c_r\right|p_n\, sin(f_{nr})\frac{1}{\hbar S}\sum\left|c_u c_v\right|p_u\, sin(f_{uv}) \qquad (xvi)$$

$$= \frac{1}{\hbar^2 S}\sum\left|c_n c_u\right|p_n p_u kos(f_{nu}) - \frac{1}{4}\left(\frac{1}{S}\frac{\partial S}{\partial x}\right)^2 \qquad (xvii)$$

Persamaan (xv) dan (xvi) memberikan bahagian pertama teorem.

Persamaan R didapati menerusi hubungan R = √S yang memberikan

$$\frac{\partial R}{\partial t} - \gamma\frac{\partial^2 R}{\partial x^2} = \frac{1}{2\sqrt{S}}\left[\frac{\partial S}{\partial t} - \gamma\frac{\partial^2 S}{\partial x^2} - \frac{\gamma}{S}\left(\frac{\partial S}{\partial x}\right)^2\right]$$

Teorem terbukti dengan menggunakan bahagian satu teorem ini dan persamaan (xvii).

KESIMPULAN

Teorem di atas membuktikan bahawa zarah bebas kuantum dapat dianggap sebagai zarah klasik yang pergerakannya dapat dimodelkan sebagai suatu resapan tidak bebas yang berpekali resapan, berpekali hanyutan dan bersumber luar W atau U dalam teorem itu. Ini tentunya boleh menguatkan lagi keesahan tafsiran Born (tafsiran piawai atau tafsiran Copenhagen) tentang tabii kebarangkalian tentang zarah kuantum itu dengan lebih bersahaja lagi kerana penyelesaian persamaan resapan inilah yang menjadi taburan kebarangkalian itu. Pada masa yang sama teorem ini juga dapat digunakan untuk menguatkan tafsiran bukan berkebarangkalian tentang mekanik kuantum seperti yang pertama kali diutarakan oleh Bohm itu, walaupun potensi kuantum kami berbeza daripada yang diperolehi Bohm itu. Kajian ini bersama dengan kajian kami sebelum ini (Shaharir dan Nik Rusdi 2000) melengkapkan program pentafsiran semula zarah bebas kuantum mengikut model resapan dan zarah bebas klasik dalam fasa pertamanya. Kajian selanjutnya ialah memodelkan hasil-hasil kajian ini menerusi kalkulus stokastik pula.

RUJUKAN

AEBI, R. 1996. *Schroedinger Diffusion Process.* Birkhauser-Verlag.

BERNDL, K., M. DAUMER, D. DURR, S. GOLDSTEIN, dan N. ZANGHI. 1995. A survey of Bohmian Mechanics. *Il Nuovo Cimento* 110B: 737-750.

BESS, L. 1994. New features of the Schroedinger diffusion model. *Physics Essays* **7**(1): 77-82.

BOHM, D. 1952. A suggested interpretation of the quantum theory in terms of "hidden variables". I and II. Phys. Rev. 85: 166-179; 180-193. Dicetak kembali dalam *Quantum Theory and Measurement*, disunting oleh Wheeler, J.A. & Zurek, W.H. 1983. Princeton Univ. Press.

___________ . 1984. *Causality and Chance in Quantum Physics.* 2nd ed. Routledge.

BOHN, D. dan B.J. HILEY. 1993. *The Undivided Universe.* London: Routledge.

BOHM, D., B.J. HILEY dan P.N. KOLOYEROU. 1987. An ontological basis for the quantum theory. *Phys. Reports* **144**(4): 321-375.

BORN, M. 1926a. On the quantum mechanics of collisions (preliminary communication). Dlm. *Quantum Theory and Measurement*, disunting oleh Wheeler, J.A. dan W.H. Zurek, 1983. Princeton Univ. Press.

___________ . 1926b. On quantum mechanics of collision. Dlm *Wave Mechanics*, disunting oleh Ludwig, G. 1968. Oxford: Pergamon Press

CARTWRIGHT, N. 1987. Max Born and the reality of quantum probability. Dalam *The Probabilistic Revolution. Vol.2: Ideas in the Sciences*, disunting oleh Kruger, L., Gigerenzer,G. & Morgan, M.S. p. 409-416. MIT Press.

CUSHING, J.T., A. FINE dan S. GOLDSTEIN. 1996. *Bohmian Mechanics and Quantum Theory: An Appraisal.* Dordrecht: Kluwer.

GILSON, J.G. 1980. A classical basis for quantum mechanics. *Ann. Ins. Henri Poincare. Sec.A: Physique Theorie* **XXXII**(4): 319-325.

___________ . 1978. The fluid process underlying quantum mechanics. *Acta Physicsa Hungaricae.* t. **44(4)**: 333

HEALEY, A. 1990. *The Philosophy of Quantum Mechanics: An Interactive Approach.* Cambridge Univ. Press.

MADELUNG, E. 1926. Quantenttheorie in hydrodynamischer form. *Zeitschrift fur Physik.* t. **40:** 332

NAGASAWA, M. 1993. *Schroedinger Equation and Diffusion Theory.* Birkhauser-Verlag.

PAIS, A. 1982. Max Born's statistical interpretation of quantum mechanics. *Science* **218** Dec: 1193-1198.

ROBINETT, R.W. 1997. *Quantum Mechanics: Classical Results, Modern System, and Visualized Examples.* N.York: OUP.

SCHIFF, I. 1968. *Quantum Mechanics.* 3rd Ed. N. York: McGraw-Hill.

SCHROEDINGER, E. 1926. Quantisation as a problem of proper values, part 1. Dalam Schroedinger, E. 1928. *Collected Papers on Wave Mechanics.* p. 1-12. London: Blackie & Sons Ltd.

SHAHARIR MOHAMAD ZAIN dan NIK RUSDI YAACOB. 2000. Zarah Schroedinger bebas sebagai suatu resapan. Laporan penyelidikan Q/3/99, Pusat Pengajian Sains Matematik, UKM.

SHARPE, K.J. 1993. *David Bohm's World: New Physics and New Religions.* London: Associated Univ. Press

SONEGO, S. 1991. Interpretation of the hydrodynamical formalism of quantum mechanics. *Foundation of Phys.* **21(10)**: 1135-1181

WALLSTROM, T.C. 1994. Inequivalence between the Schroedinger equation and the Madelung hydrodynamic equations. *Phys. Rev. A.* **49(3)**: 1613-1617

WICK, D. 1995. *The Infamous Boundry: Seven Decades of Controversy in Qantum Mechanics.* Birkhauser-Verlag.

ZAINAL ABDUL AZIZ. 1997. Penyelesaian Persamaan Resapan Kompleks yang Serasi dengan Kamiran Feynmanan. Tesis Doktor Falsafah, Jabatan Matematik, UKM. (tidak diterbitkan).

Pertanika J. Sci. & Technol. 10(1): 99 - 109 (2002)
ISSN: 0128-7680

Bahan Penderia untuk Pengesanan Aluminium Berasaskan Reagen Kolorimetrik Terdop dalam Filem Sol-Gel

Musa Ahmad & Norleen Abdul Manaf
Pusat Pengajian Sains Kimia dan Teknologi Makanan
Fakulti Sains dan Teknologi Universiti Kebangsaan Malaysia
43600 Bangi, Selangor Darul Ehsan

Diterima: 4 September 2000

ABSTRAK

Teknik sol-gel digunakan dalam kajian ini untuk menyediakan bahan penderia yang boleh digunakan untuk pengesanan aluminium dalam persekitaran air. Dua reagen organik yang sensitif terhadap aluminium iaitu eriokrom sianin R (ECR) dan krom azurol S (CAS) digunakan dalam kajian ini. Bahan mula yang digunakan untuk penyediaan larutan sol-gel ialah tetraetilortosilikat (TEOS), air dan etanol dalam nisbah isi padu TEOS:air:etanol = 30:31:30. Kaedah penyalutan celup digunakan untuk menghasilkan filem sol-gel pada permukaan penyokong kaca slaid. Hasil kajian menunjukkan bahawa reagen yang didopkan dalam filem nipis sol-gel adalah stabil secara foto, dapat mengekalkan sifat-sifat kimianya seperti dalam larutan bebas dan boleh dijana semula menggunakan larutan NaF. Kajian ini juga menunjukkan bahawa reagen terdop dalam filem nipis sol-gel masih bertindak balas dengan analit dan berpotensi untuk digunakan sebagai bahan penderia untuk pengesanan aluminium.

ABSTRACT

Sol-gel technique was used in this study for sensing material preparation which could be used for aluminium detection in aqueous environment. Two organic reagents i.e. eriochrome cyanine R (ECR) and chrome azurol S (CAS), which are sensitive to aluminium were used in this study. The precursors used for the sol-gel solution preparation were tetraethyl orthosilicate (TEOS), water and ethanol in the volume ratio of 30:31:30. Dip-coating method was used to deposit the sol-gel solution onto microscope slide glass support to produce a thin sol-gel film. The study shows that the immobilised reagents have a good photo-stability, still maintaining their chemical properties as in free solution and regenerable by using NaF solution. The study also indicates that the reagen doped in sol-gel film still react with the analyte and could be potentially used as sensing material for aluminium detection.

Keywords: Filem sol-gel, bahan penderia, mengesan aluminium

PENGENALAN

Sol-gel adalah suatu bahan oksida inorganik yang komposisi kimianya serupa dengan kaca dan seramik (Avnir 1994). Ia boleh disediakan dengan mudah pada suhu bilik dan mempunyai kehomogenan yang tinggi. Sol-gel bersifat lengai secara kimia dan stabil secara terma. Sifat filem sol-gel yang lutsinar dan poros menyebabkan ia sesuai digunakan untuk penyediaan bahan penderia semasa pembinaan penderia kimia berasaskan gentian optik (Brinker

& Scherer 1990). Sifat poros bahan $(SiO_mH_n)_p$ yang diguna membolehkan analit membaur dalam taburan liang-liang mikroskopik yang mengandungi reagen terpegun yang membentuk suatu luas permukaan yang tinggi.

Filem nipis sol-gel telah banyak digunakan sebagai fasa reagen untuk penderia gentian optik. Dalam penyelidikan ini, filem nipis sol-gel digunakan sebagai matrik untuk tindak balas antara aluminium dengan reagen kolorimetri ECR dan CAS. Dengan kehadiran aluminium, perubahan warna akan berlaku terhadap reagen terdop kerana pembentukan kompleks antara analit aluminium dengan reagen terdop berkenaan. Perubahan warna ini dapat dipantau menggunakan gentian optik dan fenomena ini boleh digunakan sebagai asas untuk pembinaan penderia kimia aluminium menggunakan gentian optik.

EKSPERIMEN

Bahan-bahan Kimia

Bahan-bahan kimia yang digunakan dalam kajian ini adalah bahan pemula sol-gel seperti tetraetilortosilikat, TEOS (Fluka Chemika, >>99.0 %); triton X-100 (Fluka AG) dan etanol (BDH Spectrograde, 99.7%). Bahan-bahan kimia lain yang digunakan termasuklah asid hidroklorik, HCl (Riedel-deHaen); natrium hidroksida, NaOH (J.T Baker Inc, 98.0%); asid asetik glasial (BDH, 99.5%); natrium asetat (Aldrich); natrium florida (Fluka) dan aluminium (III) nitrat (Fluka, 98.0 %). Reagen organik yang digunakan adalah ECR (Searle Company) dan CAS (Fluka Chemika). Air yang digunakan dalam kajian ini adalah air nyah ion.

Peralatan

Spektrofotometer UL-Nampak model Shimadzu 160A yang menggunakan sumber lampu tungsten dan deutrium digunakan untuk pengukuran semua serapan dalam kajian ini.

Cara Kerja

Reagen. Larutan stok reagen ECR, 5.0 x 10^{-2} M dan larutan stok reagen CAS, 2.0 x 10^{-2} M masing-masingnya disediakan dengan melarutkan 2.68 g reagen ECR dan 1.21 g reagen CAS dalam air dan menjadikan larutan ke isipadu 100ml dengan larutan larutan penimbal. Larutan stok aluminium, 2.0 x 10^{-2}M disediakan dengan melarutkan 0.75g $Al(NO_3)_3$ dalam air dan menjadikan larutan ke 100 ml dengan larutan penimbal.

Larutan sol-gel disediakan dengan mencampurkan tetraetilortosilikat (TEOS), etanol dan air nyah ion dalam nisbah 30 : 31 : 30 mengikut isi padu. Sebanyak 5 ml surfaktan triton X-100 dan beberapa titik asid hidroklorik pekat kemudiannya ditambahkan ke dalam campuran dan larutan ini dikacau dengan pengacau magnet selama 2 jam.

Penyediaan lapisan filem nipis. Kaedah penyalutan celup digunakan dalam kajian ini untuk penyediaan filem sol-gel dan kaca slaid digunakan sebagai penyokong. Penyokong dibersihkan terlebih dahulu dengan etanol dan dipastikan dalam keadaan kering sebelum digunakan. Penyalutan celup dilakukan dengan mencelupkan penyokong kaca slaid ke dalam larutan sol-gel yang mengandungi reagen dan ditarik dengan perlahan bagi memperolehi lapisan yang nipis. Penyokong ini kemudiannya dibiarkan kering dalam desikator dalam keadaan menegak.

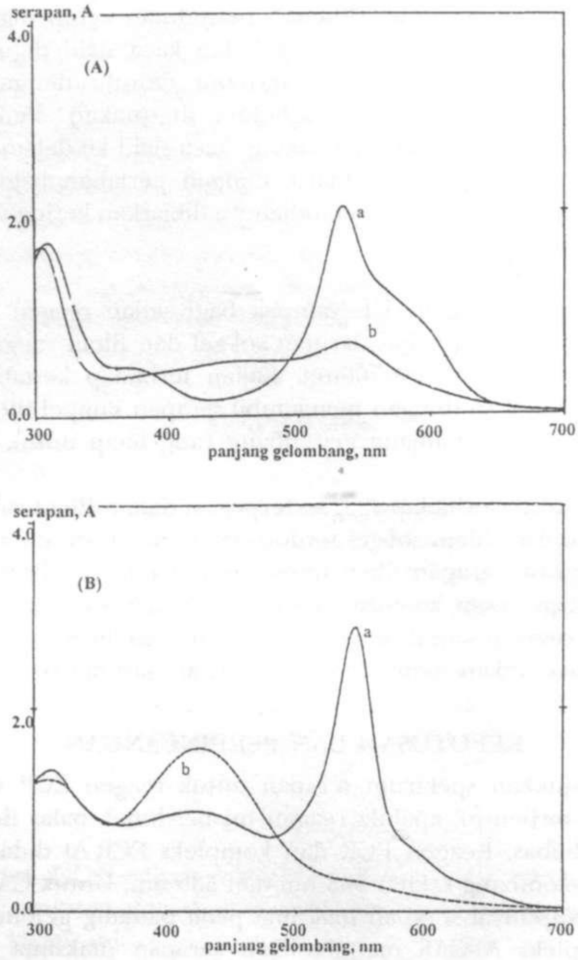
Pencirian. Spektrum serapan UL-Nampak bagi setiap reagen dan kompleks aluminium dalam larutan bebas, larutan sol-gel dan filem sol-gel diukur pada julat panjang gelombang 300–700nm. Kajian terhadap kestabilan foto bagi reagen terdop dilakukan dengan mengambil serapan sampel filem nipis sol-gel terdop berkenaan pada panjang gelombang yang tetap untuk setiap 5 minit selama 8 jam.

Kajian kelunturan terhadap reagen terpegun dalam filem sol-gel dilakukan dengan mencelupkan filem sol-gel terdop dalam air nyah ion selama 5 minit. Setelah dikeringkan, serapan filem nipis tersebut diambil. Proses ini diulangi sehingga tiada lagi kesan kelunturan yang diperhatikan. Kemungkinan filem sol-gel terdop dijanakan semula setelah digunakan juga dikaji dalam penyelidikan ini dengan mencelupkan filem tersebut ke dalam larutan natrium florida.

KEPUTUSAN DAN PERBINCANGAN

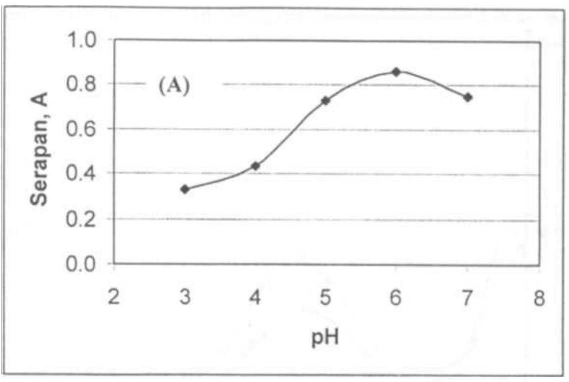
Rajah 1 menunjukkan spektrum serapan untuk reagen ECR dan CAS serta kompleks yang terbentuk apabila reagen ini bertindak balas dengan ion Al^{3+} dalam larutan bebas. Reagen ECR dan kompleks ECR-Al didapati menyerap pada panjang gelombang sekitar 538 nm dan 536 nm. Untuk CAS pula, reagen ini didapati mempunyai serapan maksima pada panjang gelombang 426.6nm manakala kompleks Al-CAS menghasilkan serapan maksima pada panjang gelombang 546.4nm. Nilai ini hampir menyamai nilai panjang gelombang maksima yang dilaporkan dalam literatur iaitu 535nm (Ariza & Gonzalez 1984; Corbett & Guerin 1966), dan 536nm (Sampson & Fleck 1984) untuk kompleks ECR-Al serta 546nm (Kennedy & Powell 1986) dan 567nm (Pakalns 1965) untuk kompleks CAS-Al.

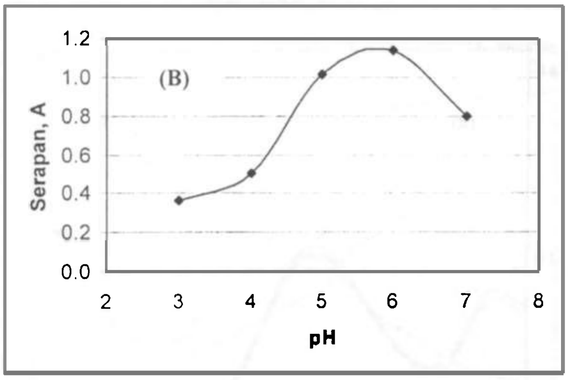
Rajah 2 menunjukkan kesan pH terhadap pembentukan kompleks aluminium dengan reagen ECR dan CAS. Seperti yang ditunjukkan, medium pH yang paling sesuai untuk pembentukan kompleks ECR-Al dan CAS-Al adalah pH 6.0. Pada pH yang terlalu asid, ion-ion Al^{3+} cenderung berada dalam keadaan ion manakala pada pH yang terlalu bes, ion-ion Al^{3+} akan membentuk kompleks hidroksida. Corbett dan Guerin (1966) melaporkan bahawa nilai pH optimum untuk tindak balas pembentukan kompleks ECR-Al adalah pada pH 6.1 manakala untuk pembentukan kompleks CAS-Al pula, Pakalns (1965) melaporkan bahawa pH optimum yang diperlukan adalah 5.8.



Rajah 1: Spektrum serapan untuk reagen ECR (A) dan CAS (B) dalam larutan bebas (b) dan kompleks A1 (b) dalam medium yang sama. Kepekatan reagen dan A1 adalah masing-masingnya 6.4 × 10^{-4} M dan 1.6 × 10^{-4}

Spektrum serapan untuk reagen ECR dan CAS serta kompleksnya dengan Al^{3+} dalam larutan sol-gel ditunjukkan dalam *Rajah 3*. Membandingkan spektrum serapan untuk reagen dan kompleks reagen-Al dalam larutan bebas dan dalam larutan sol-gel, kedua-dua reagen dan kompleks reagen-Al didapati menyerap pada panjang gelombang yang sama dalam larutan sol-gel. Reagen ECR dan kompleks ECR-Al didapati menyerap pada panjang gelombang 470.0nm manakala reagen CAS dan kompleks CAS-Al didapati menyerap pada panjang gelombang 460.0nm. Selain itu, perbezaan keamatan serapan antara reagen dan kompleks reagen-Al didapati terlalu rendah. Keadaan ini berlaku disebabkan oleh pH medium larutan sol gel yang tinggi iaitu 1.5. Nilai pH larutan yang terlalu berasid didapati tidak sesuai bagi pembentukan kompleks aluminium



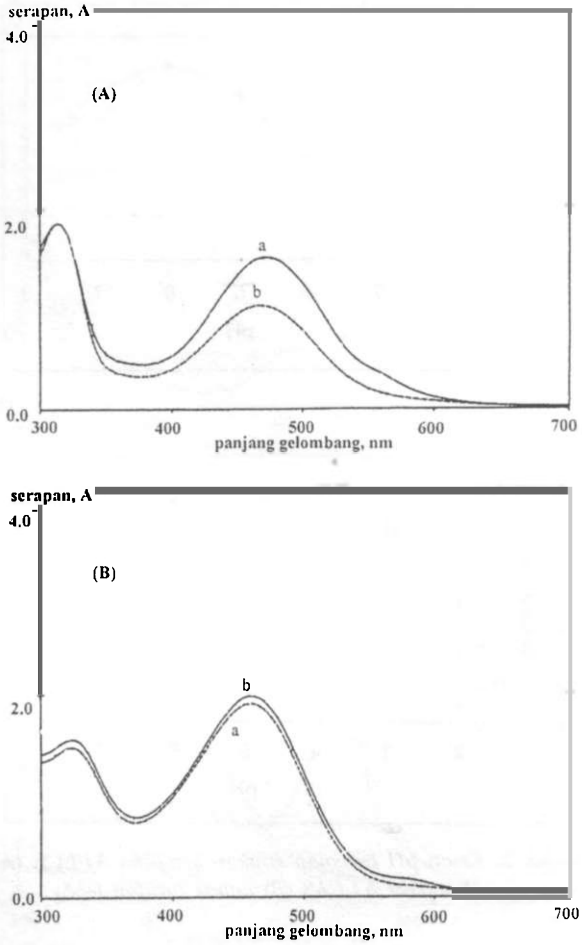


Rajah 2: Kesan pH terhadap serapan kompleks Al-ECR (A) dan kompleks Al-CAS (B) dalam larutan bebas

dan ini menyebabkan nilai keamatan serapan untuk kompleks aluminium terlalu kecil. Seperti ditunjukkan dalam *Rajah 2*, medium pH yang paling sesuai untuk pembentukan kompleks ECR-Al dan CAS-Al adalah pH 6.0.

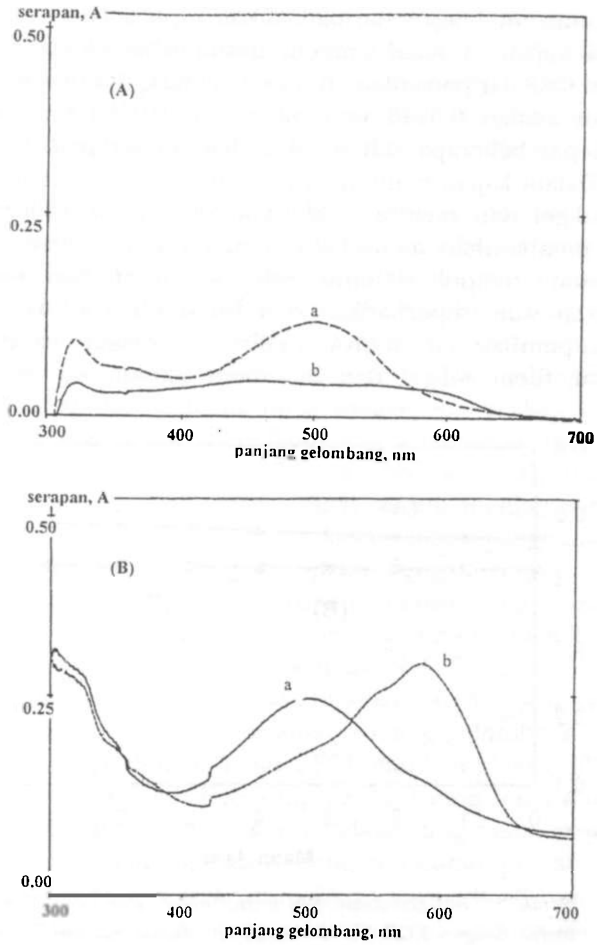
Bagi filem nipis sol gel terdop, warna reagen dan kompleks Al^{3+} yang terpegun adalah sama seperti yang didapati dalam larutan bebas. Walau bagaimanapun dengan menggunakan kepekatan reagen dan Al^{3+} yang sama seperti yang digunakan dalam larutan bebas dan larutan sol gel, serapan untuk reagen dan kompleks reagen-Al dalam filem nipis sol-gel tidak dapat dicerap kerana serapan yang diperolehi adalah amat rendah. Ini disebabkan oleh kuantiti reagen yang terperangkap dalam jaringan liang-liang dalam filem nipis sol-gel adalah kecil. Oleh itu kompleks Al^{3+} yang terbentuk juga sedikit. Faktor yang menyebabkan serapan reagen terdop yang rendah bagi filem nipis ini dinamakan kesan pemerangkapan. Kesan ini menyebabkan interaksi antara analit dengan molekul terdop rendah kerana molekul terdop tidak bebas bergerak.

Untuk membolehkan spektrum serapan untuk reagen dan kompleksnya dengan Al^{3+} diperolehi, kepekatan reagen dan Al^{3+} yang lebih pekat telah



Rajah 3: Spektrum serapan untuk reagen ECR (A) dan CAS (B) dalam larutan sol-gel (a) dan kompleks Al dalam medium yang sama (b). Kepekatan reagren dan Al adalah masing-masing 6.4×10^{-4} *M dan* 1.6×10^{-4} *M*

digunakan dalam kajian ini. *Rajah 4* menunjukkan spektrum serapan filem nipis sol-gel yang telah didopkan masing-masing dengan reagen ECR dan CAS serta kompleks Al^{3+}. Spektrum tersebut menunjukkan bahawa serapan untuk reagen dan kompleks reagen-Al dalam filem nipis sol-gel berlaku pada panjang gelombang yang lebih tinggi daripada panjang gelombang serapan reagen dan kompleks Al^{3+} masing-masing dalam larutan bebas. Untuk surfaktan ionik, kesan anjakan ini telah diterangkan dalam literatur disebabkan oleh interaksi spesifik antara reagen terpegun dan agen aktif-permukaan yang digunakan. Oleh kerana surfaktan yang digunakan dalam kajian ini adalah jenis non-ionik, anjakan pada panjang gelombang serapan maksima reagen terpegun dijangka

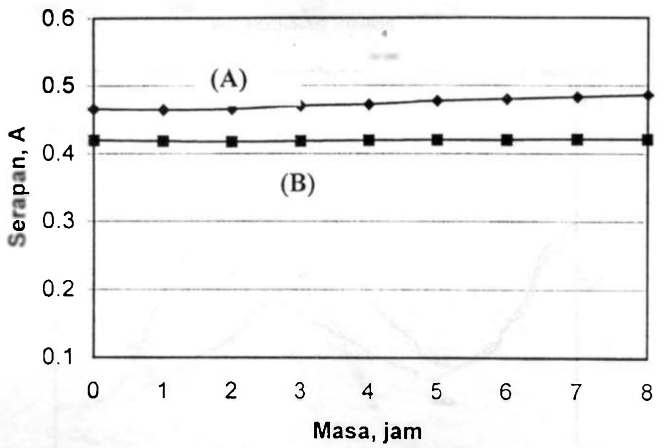


Rajah 4: Spektrum serapan untuk reagen ECR (A) dan CAS (B) dalam filem sol-gel (a) dan kompleks Al dalam medium yang sama (b)

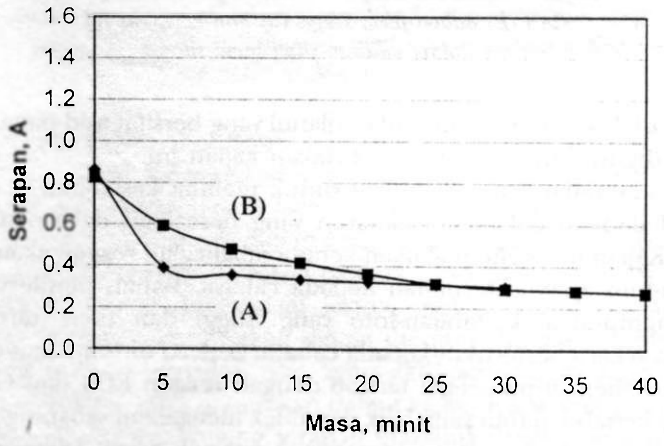
disebabkan oleh kehadiran kumpulan silanol yang bersifat asid pada permukaan silika yang digunakan sebagai matrik dalam kajian ini.

Analisis kestabilan foto dilakukan untuk melihat kestabilan reagen terdop apabila didedahkan dalam persekitaran yang bercahaya dalam suatu tempoh yang lama. Kajian ini perlu dilakukan kerana sebahagian reagen akan mengalami penguraian-foto apabila terdedah kepada cahaya. Bahan penderia yang baik mestilah mempunyai kestabilan-foto yang tinggi dan tidak terurai apabila didedahkan secara berterusan kepada cahaya. Seperti ditunjukkan dalam *Rajah 5*, kedua-dua filem nipis sol-gel terdop dengan reagen ECR dan CAS didapati mempunyai kestabilan-foto yang baik dan tidak mengalami sebarang penguraian-foto yang nyata, sekurang-kurangnya untuk tempoh 8 jam kajian dilakukan.

Analisis kelunturan reagen daripada filem nipis sol-gel terdop juga telah dilakukan dalam kajian ini. *Rajah 6* menunjukkan bahawa kadar awal kelunturan reagen ECR dan CAS daripada filem nipis sol gel yang disediakan menggunakan kaedah celupan adalah 0.0146 unit/minit dan 0.0199 unit/minit, masing-masingnya. Selepas beberapa kali basuhan bacaan serapan didapati menjadi hampir tetap. Dalam kajian bandingan penyediaan bahan penderia untuk pH dalam filem sol-gel dan membran kitosan, Musa *et al.* (2001) melaporkan bahawa kesan pembasuhan menyebabkan pelunturan reagen daripada filem sol-gel untuk suatu tempoh tertentu. Selepas tempoh berkenaan, tiada lagi kesan pelunturan yang diperhatikan dan ini disebabkan kesan pelunturan semasa proses pembasuhan hanya melibatkan reagen yang terpegun di permukaan luar filem sol-gel dan membran kitosan sahaja. Reagen yang



Rajah 5: Graf kestabilan foto untuk filem nipis sol-gel yang terdop dengan ECR (A) dan CAS (B) dalam tempoh 8 jam



Rajah 6: Graf kelunturan reagen (A) ECR dan (B) CAS yang didopkan dalam filem nipis sol-gel yang disediakan mengguna kaedah celupan

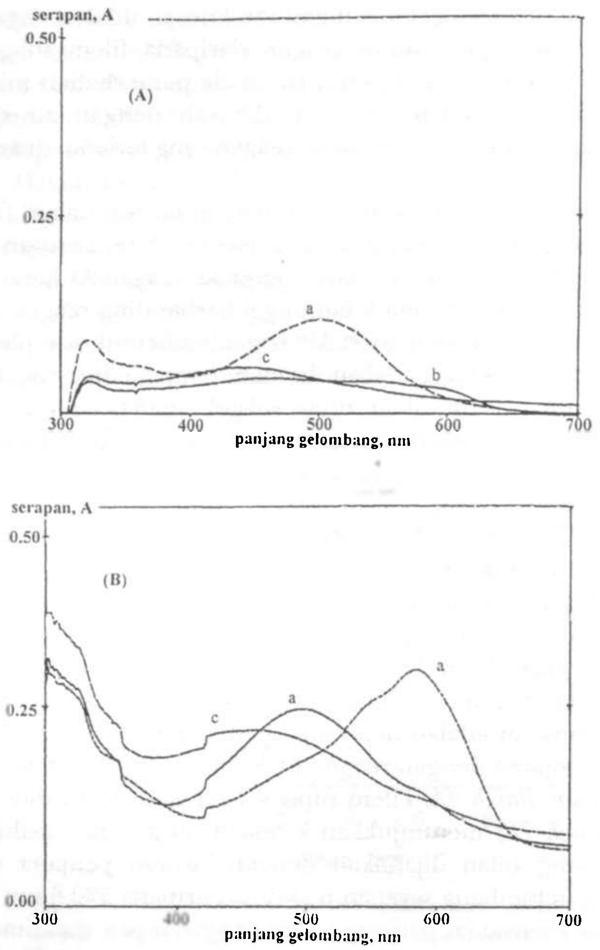
terpegun di dalam filem sol-gel dan membran kitosan tidak mengalami masalah keluntūran ini. Corak peluntūran reagen daripada filem sol-gel juga telah dilaporkan oleh Rottman *et al.* (1992). Daripada pemerhatian ini, filem sol-gel yang digunakan dalam kajian ini akan dibasuh dengan air suling dahulu sebelum digunakan untuk menanggalkan reagen yang terjerap di atas permukaan filem sol-gel.

Setelah bertindak balas dengan analit, reagen terdop dalam filem nipis sol-gel boleh dijanakan semula menggunakan larutan NaF. Larutan ini berfungsi untuk mengeluarkan ion-ion Al^{3+} dari kompleks reagen-Al kerana ion florida mempunyai kecenderungan yang lebih tinggi berbanding reagen ECR dan CAS untuk bertindak balas dengan ion-ion Al^{3+} bagi membentuk kompleks aluminium florida, AlF_3. Dengan menggunakan larutan penjana ini, spektrum serapan kompleks reagen-Al dalam filem nipis sol-gel dijangkakan akan menyamai spektrum serapan reagen asal dalam filem sol-gel setelah filem terdop berkenaan dicelupkan dalam larutan penjana. Larutan NaF telah digunakan oleh Musa dan Narayanaswamy (1995a, 1995b, 1994) untuk menjanakan semula prob gentian optik aluminium setelah ia digunakan dalam pengesanan ion-ion Al^{3+}.

Bagi filem nipis sol-gel yang didopkan dengan reagen ECR, *Rajah 7A* menunjukkan bahawa terdapat penurunan intensiti yang tinggi bagi serapan filem nipis sol-gel yang didopkan dengan kompleks ECR-Al, setelah filem berkenaan dicelupkan dalam larutan penjana, NaF. Selepas penjanaan, panjang gelombang serapan maksimum didapati teranjak dari 502.4nm kepada 440.0nm. Panjang gelombang ini adalah menyamai panjang gelombang untuk filem sol-gel yang telah didopkan dengan reagen ECR dan kompleks ECR-Al seperti yang ditunjukkan dalam *Rajah 4A*. Filem nipis sol-gel yang telah didopkan dengan reagen CAS (*Rajah 7B*) menunjukkan keadaan yang sama, apabila filem nipis sol-gel terdop yang telah dijanakan dengan larutan penjana menunjukkan anjakan panjang gelombang serapan maksima daripada 580.6 nm kepada 460.0 nm iaitu hampir menyamai panjang gelombang serapan maksima untuk filem sol-gel yang telah didopkan dengan reagen CAS dan kompleks CAS-Al seperti ditunjukkan dalam *Rajah 4B*. Keadaan ini berlaku kerana setelah penjanaan dilakukan, ion-ion Al^{3+} akan dikeluarkan dan apa yang tinggal dalam filem sol-gel hanyalah reagen yang sedia untuk digunakan semula dalam pengesanan ion-ion Al^{3+} yang seterusnya.

KESIMPULAN

Hasil penyelidikan ini menunjukkan bahawa filem nipis sol-gel merupakan suatu matrik yang baik untuk penyediaan bahan penderia bagi pengesanan Al menggunakan reagen ECR dan CAS. Serapan filem nipis sol-gel bagi reagen dan kompleks Al^{3+} yang diperolehi, berlaku pada panjang gelombang yang hampir sama seperti dalam larutan bebas. Ini menunjukkan bahawa filem nipis sol-gel terdop merupakan suatu substrat yang baik bagi tindak balas dengan Al^{3+} dan ia dapat digunakan sebagai fasa reagen dalam pembinaan penderia kimia aluminium berasaskan gentian optik. Filem nipis sol-gel terdop juga menunjukkan kestabilan foto yang agak baik apabila didedahkan dalam



Rajah 7: Spektrum serapan filem nipis sol-gel terdop dengan ECR (A) dan CAS (B) sebelum (a) dan selepas bertindak balas dengan aluminium (b) dan setelah dijanakan semula dengan larutan natrium florida (a)

persekitaran yang bercahaya dalam tempoh yang lama dan mempunyai kadar pelunturan yang rendah. Bahan penderia yang disediakan dalam kajian ini didapati boleh dijanakan semula menggunakan larutan penjana NaF. Serapan filem nipis sol gel yang dijanakan semula berlaku pada panjang gelombang yang hampir dengan panjang gelombang filem nipis yang terdop dengan reagen asal.

PENGHARGAAN

Penyelidik ingin merakamkan penghargaan kepada Kementerian Sains, Teknologi & Alam Sekitar kerana sumbangan gran penyelidikan IRPA 03-02-02-0044 untuk kajian ini.

RUJUKAN

ARIZA J.L.G. dan M.L.M. GONZALEZ. 1984. N,N'-Oxalylbis(salicylaldehydehydrazone) as an Analytical Spectrophotometric and Fluorimetric Reagent: Part 1 – Study of the metal reactivity and application to the determination of aluminium. *Analyst* 198-**885**.

AVNIR, D. 1994. Organic chemistry within ceramic matrices: doped sol gel materials. *Acc. Chem. Res.* **28**: 328-334.

BRINKER, C.J. dan G.W. SCHERER. 1990. *Sol Gel Science: The Physics & Chemistry of Sol Gel Processing.* Boston: Academic Press Inc.

CORBETT J.A. dan B.D. GUERIN. 1966. The determination of aluminium in iron and steel. *Analyst* **9**: 490.

KENNEDY J.A. dan H.K.J. POWELL. 1986. Colorimetric determination of aluminium (III) with chrome azurol S and the reactivity of hydrolysed aluminium species. *Anal. Chim. Acta*: 184-329.

MUSA AHMAD, NOREZUNY MOHAMAD dan JARIAH ABDULLAH. 2001. Sensing material for oxygen gas prepared by doping sol-gel film with Tris(2,2-bipyridyl)dichlororuthenium complex. *J. Non-Crystalline Solids* **290**: 86-91.

MUSA AHMAD dan RAMAIER NARAYANASWAMY. 1995a. A flow-cell optosensor for monitoring aluminium (III) based on immobilised eriochrome cyanine R and reflectance spectrophotometry. *The Science of the Total Environment* **163**: 221-227 .

MUSA AHMAD dan RAMAIER NARAYANASWAMY. 1995b. Development of optical fibre Al(III) sensor based on immobilised chrome azurol S. *Talanta* **42**:1337- 1344.

MUSA AHMAD dan RAMAIER NARAYANASWAMY 1994, Fibre optic reflectance sensor for the determination of aluminium (III) in aqueous environment. *Anal. Chim. Acta* **291**:255 - 260.

PAKALNS P. 1965. Spectrophotometric determination of aluminium with chrome azurol S. *Anal. Chim. Acta* **32**: 57.

ROTTMAN C., M. OTTOLENGHI, R.ZUSMAN, O.LEV, M. SMITH, G. GONG, M.L. KAGAN dan D. AVNIR. 1992. *Material Letters* **13**: 293.

SAMPSON B. dan A. FLECK. 1984. Measurement of aluminium in dialysis fluid and water by spectrophotometric procedure. *Analyst* **109**: 369.

Pertanika J. Sci. & Technol. **10(1)**: 111 - 118 (2002)

ISSN: 0128-7680

On Solutions of Some Polynomial-Functional Equations

U.D. Bekbaev
Department of Mathematics
Faculty of Science and Environmental Studies
Universiti Putra Malaysia
43400 UPM Serdang, Selangor, Malaysia

Received: 16 September 1998

ABSTRAK

Sistem persamaan polinomial-fungsian tertentu yang dijanakan daripada sifat-sifat biasa pembezaan diperhatikan. Penyelesaian sistem ini ke atas medan dan bentuk kanonik mereka relatif terhadap kesetaraan natural diberikan.

ABSTRACT

Systems of some polynomial-functional equations, which are derived from the usual properties of derivative, are considered. Solutions of these systems over a field and their canonical forms relative to natural equivalences are given.

Keywords: Algebra, derivative, polynomial

INTRODUCTION

Let R be a commutative, associative algebra over a field F of characteristic zero, $d : R \rightarrow R$ be a derivative i.e. d is an additive map and $d(ab) = d(a)b+ ad(b)$ for any $a,b \in R$. The last equality means that to know the derivative of a product it is enough to know the derivatives of the factors because of polynomialness $d(ab)$ in a, b, $d(a)$ and $d(b)$ over F. If for a while we denote $d(xy)$ as a polynomial $f[x, y, d(x), d(y)]$ then the equality $d((xy)z)= d(x(yz))$ draws in the equality

$$f[xy, z, f[x, y, d(x), d(y)], d(z)]=f[x, yz, d(x), f[y, z, d(y), d(z)]],$$

the equality $d(xy) = d(yx)$ draws in the equality

$$f[x, y, d(x), d(y)] = f[y, x, d(y), d(x)]$$

Therefore the following problem is natural: Let x_1,x_2,x_3,x_4,x_5,x_6 be indeterminates over the field F. Find all such polinomials $f[x_1,x_2,x_3,x_4]$ over F for which the following system of polynomial-functional equations is valid.

$$\begin{cases} f[x_1x_2, x_3, f[x_1, x_2, x_4, x_5], x_6] = f[x_1, x_2x_3, x_4, f[x_2, x_3, x_5, x_6]] \\ f[x_1x_2, x_3, x_4] = f[x_2, x_1, x_4, x_3] \end{cases} \quad (1)$$

The paper deals with this problem. To follow the proofs of the results below one needs only some basic notions on polynomials, which can be found nearly in any text book on algebra e.g. in (Waerden 1991).

Lemma. The first polynomial-functional equation of system (1) has only the following solutions in $F[x_1,x_2,x_3,x_4]$:

$$f = \sum_{i=0}^{k} a_i (x_1x_2)^i \text{ ,} where\ k \in N \cup \{0\}\ and\ a_i \in F;$$

$$f = x_1^n x_4 + \sum_{i=0}^{k} a_i (x_1^n - x_1^i) x_2^i, \ where\ k,\ n \in N \cup \{0\} \text{ and } a_i \in F;$$

$$f = x_2^n x_3 + \sum_{i=0}^{k} a_i x_1^i (x_2^n - x_2^i) x_2^i \ where\ k,\ n \in N \cup \{0\} \text{ and } a_i \in F;$$

$$f = x_1^n x_4 + x_2^n x_3 + a x_1^m x_2^n, \ where\ m,n \in N \cup \{0\},\ m \neq n\ and\ a \in F;$$

$$f = x_1^n x_4 + x_2^n x_3 + \sum_{i=0}^{k} a_i (x_1^i x_2^n + x_1^n x_2^i - x_1^i x_2^i),$$

where $k,\ n \in N \cup \{0\}$ and $a_i \in F$;

$$f = g[x_1]x_4 + g[x_2]x_3 + a x_3 x_4 + \frac{1}{a}(g[x_1]g[x_2] - g[x_1x_2] - g[x_1x_2])$$

where $g[x_1] \in F[x_1]$ *and* $a \in F^* = F \setminus \{0\}$.

Proof of lemma. Let $f \in F[x_1,x_2,x_3,x_4]$ be a polynomial for which the first equality of (1) is valid. Comparing degrees on the left and the right sides of this equality in x_4 (x_6) as polynomials over $F[x_1,x_2,x_3,x_5,x_6]$ (corresp. $F[x_1,x_2,x_3,x_4,x_5]$) one has inequality

$$\deg_{x_3} f[x_1,x_2,x_3,x_4] \leq 1 \text{ (corresp. } \deg_{x_4} f[x_1,x_2,x_3,x_4] \leq 1)$$

Due to this

$$f[x_1,x_2,x_3,x_4] = x_3x_4 f_0[x_1,x_2] + x_3 f_1[x_1,x_2] + x_4 f_2[x_1,x_2] + f_3[x_1,x_2], \tag{2}$$

where $f_i[x_1,x_2] \in F[x_1, x_2]$. Considering the first equality of (1) as an equality of polynomials in x_4, x_5, x_6 over $F[x_1,x_2,x_3]$ it can be written as the following system of equalities (3)

$$\left|\begin{array}{l}
f[x_1,x_2]f_0[x_1x_2,x_3] = f_0[x_2,x_3]f_0[x_1,x_2x_3] \\
f_1[x_1x_2]f_0[x_1x_2,x_3] = f_2[x_2,x_3]f_0[x_1,x_2x_3] \\
f_2[x_1,x_2]f_0[x_1x_2,x_3] = f_0[x_2,x_3]f_2[x_1,x_2x_3] \\
f_0[x_1,x_2]f_1[x_1x_2,x_3] = f_0[x_1,x_2x_3]f_1[x_2,x_3] \\
f_1[x_1,x_2]f_1[x_1x_2,x_3] = f_0[x_1,x_2x_3]f_3[x_2,x_3] + f_1[x_1,x_2x_3] \\
f_1[x_1x_2,x_3]f_2[x_1,x_2] = f_1[x_2,x_3]f_2[x_1,x_2x_3] \\
f_0[x_1x_2,x_3]f_3[x_1,x_2] + f_2[x_1x_2,x_3] = f_2[x_2,x_3]f_2[x_1,x_2x_3] \\
f_3[x_1,x_2]f_1[x_1x_2,x_3] + f_3[x_1x_2,x_3] = f_2[x_1,x_2x_3]f_3[x_2,x_3] + f_3[x_1,x_2x_3]
\end{array}\right.$$

Comparing degrees on the left and the right sides of the first equality of this system in x_1 and x_3 one gets $f_0 \in F$.

Let us consider first the $f_0 = 0$ case. In this case the 5th (7th) equality of the above system shows that $f_2[x_1, x_2]$ (corresp. $f_2[x_1, x_2]$) does not depend on x_1 (corresp.x_2) i.e. $f_1[x_1, x_2]=f_1[x_2]$ (corresp. $f_2[x_1,x_2]=f_2[x_1]$). Moreover $f_1[x_2]= ax_2^n$ ({corresp.$f_2[x_1]= bx_1^m$), where a (corresp. b) is 0 or 1 and n (corresp. m) $\in N \cup\{0\}$. In this case system (3) reduces to the equality

$$ax_3^n f_3[x_1,x_2]+ f_3[x_1x_2, x_3] = bx_1^m f_3[x_2,x_3] + f_3[x_1,x_2x_3] \tag{4}$$

If $a = b = 0$ then equation (4) has only the following solutions

$$f_3[x_1,x_2] = \sum_{i=0}^{k} a(x_1x_2)^i \text{ where } k \in \mathrm{N} \cup \{0\} \text{ and } a_i \in F$$

and therefore in this case due to (2) one has

$$f = f_3[x_1,x_2] \sum_{i=0}^{k} a(x_1x_2)^i, \text{ where } k \in N \cup \{0\} \text{ and } a_i \in F.$$

If $a = 1, b = 0$ then equation (4) has only the following solutions

$$f_3[x_1,x_2] = \sum_{i=0}^{k} a_i x_1^i(x_2^n - x_2^i), \text{ where } k \in N \cup\{0\} \text{ and } a_i \in F.$$

and therefore in this case due to (2) one has

$$f = x_2^n x_3 + \sum_{i=0}^{k} a_i x_1^i(x_2^n - x_2^i), \text{ where } k,\ n \in N \cup \{0\}\ a_i \in F.$$

If a = 0, b = 1 then equation (4) has only the following solutions

$$f_3[x_1,x_2] = \sum_{i=0}^{k} a_i(x_1^m - x_1^i)x_2^i, \text{ where } k_\in \mathrm{NU}\{0\} \text{ and } a_{i\in}\mathrm{F}^1$$

and therefore in this case due to (2) one has

$$f = x_1^m x_4 + \sum_{i=0}^{k} a_i(x_1^m - x_1^i)x_2^i, \text{ where K, } m_\in \mathrm{NU}\{0\} \text{ and } a_i\ \mathrm{F}^1$$

$$f_3[x_1,x_2] = \sum_{i=0}^{k} a_i(x_1^m - x_1^i)x_2, \text{ where}$$

If $a = b = 1$ and $n \neq m$ then equation (4) has only the following solutions

$$f_3[x_1,x_2] = cx_1^m x_2^n, \text{ where } c \in F.$$

and therefore in this case due to (2) one has

$$f = x_1^m x_4 + x_2^n x_3 + cx_1^m x_2^n, \text{ where } m,\ n \in N \cup \{0\}, \text{ and } c \in F.$$

If $a = b = 1$ and $n = m$ then equation (4) has only the following solutions

$$f_3[x_1,x_2] = \sum_{i=0}^{k} a_i(x_1^i x_2^n + x_1^n x_2^i - x_1^i x_2^i),$$

where $k \in N \cup \{0\}$ and $a_i \in F$ and therefore in this case due to (2) one has

$$f = x_1^n x_4 + x_2^n x^3 + \sum_{i=0}^{k} a_i(x_1^i x_2^n + x_1^n x_2^i - x_1^i x_2^i),$$

where $k,n \in N \cup \{0\}$ and $a_i \in F$.

Let us now consider the $f_0 \neq 0$ case. In this case the second equality of system (3) shows that $f_1[x_1,x_2] = f_2[x_2,x_3]$ i.e. $g[x_2] = f_1[x_1,x_2] = f_2[x_2,x_3]$. Moreover due to the 5th equality of (3) one has

$$f_3[x_2,\ x_3] = \frac{1}{f_0}\ (g[x_2]g[x_3] - g[x_2x_3])$$

and after this fact the other equalities of (3) become identities. Thus in this case one has

$$f = g[x_1]x_4 + g[x_2]x_3 + ax_3x_4 + \frac{1}{a}\ (g[x_1]g[x_2] - g[x_1x_2]), \text{ where } g[x] \in F[x] \text{ and}$$

$a \in F^*$. The proof of lemma is completed.

Now it is easy to list all solutions of system (1). To do this it is enough to choose only those polynomials from the list presented in the lemma for which the equality

$$f\,[x_1,\ x_2,\ x_3,\ x_4] = f\,[x_2,\ x_1,\ x_4,\ x_3]$$

holds. Let us present the result as the following theorem.

Theorem 1.

System (1) has only the following solutions in $F[x_1, x_2, x_3, x_4]$:

$$f = \sum_{i=0}^{k} a(x_1 x_2)^i, \textit{ where } k \in N \cup \{0\} \text{ and } a_i \in F,$$

$$f = x_1^n x_4 + x_2^n x_3 + \sum_{i=0}^{k} a_i (x_1^i x_2^n + x_1^n x_2^i - x_1^i x_2^i),$$

where $k, n \in N \cup \{0\}$ and $a_i \in F$,

$$f = g[x_1]x_4 + g[x_2]x_3 + a x_3 x_4 + \frac{1}{a}(g[x_1]g[x_2] - g[x_1 x_2]),$$

where $g[x] \in F[x]$ and $a \in F^*$.

One can check by direct calculation or use Theorem 1 be sure that if f is a solution for system (1) then

$$\bar{f}[x_1, x_2, x_3, x_4] = af[x_1, x_2, a^{-1}(x_3 - p[x_1]), a^{-1}(x_4 - p[x_2])] + p[x_1 x_2]]$$

is also solution for system (1)for any $a \in F^*$ and $p[x] \in F[x]$. In other words the set of all solutions of system (1) is invariant with respect to the following action of the group $G_1 = F^* \ltimes F[x]$ on it:

$$((a,p[x]), f[x_1, x_2, x_3, x_4]) \mapsto a f[x_1, x_2, a^{-1}(x_3 - p[x_1]), a^{-1}(x_4 - p[x_2])] + p[x_1 x_2]$$

It correspondence to the fact that if $d: R \to R$ is such an operation that $d(xy)$ is a polynomial in x, y, dx, dy over F and $a \in F^*$, $p[x] \in F[x]$ then $\delta(xy)$ is also a polynomial in x, y, δx, δy over F, where $\delta = ad + p$ i.e. $\delta(x) = ad(x) + p[x]$ by definition.

The following result gives the "simplest" forms of $d(x_1 x_2)$ with respect to that action.

Theorem 1'.

With an accuracy of the above action of group G_1 there are only the following equalities:

1) $d(x_1 x_2) = o$, which corresponds to $f = 0$,
2) $d(x_1 x_2) = d(x_1)x_2^n + x_1^n d(x_2)$, where n in $N \cup \{0\}$, which corresponds to $f = x_3 x_2^n + x_1^n x_4$,
3) $d(x_1 x_2) = d(x_1)d(x_2)$, which corresponds to $f = x_3 x_4$.

Proof of Theorem 1'. If

$$d(x_1x_2) = \sum_{i=0}^{k} a_i(x_1x_2)^i, \text{ where } k \in N \cup \{0\} \text{ and } a_i \in F$$

then for $\delta = d + p$, where $p[x] = \sum_{i=0}^{k} a_i x^i$ one has $\delta(x_1x_2) = o$.

If

$$d(x_1x_2) = d(x_1)x_2^n + x_1^n d(x_2) + \sum_{i=0}^{k} a_i(x_1^i x_2^n + x^n x_2^i - x_1^i x_2^i),$$

where $k, n \in N \cup \{0\}$ and $a_i \in F$, then for $\delta = d + p$, where $p[x] = \sum_{i=0}^{k} a_i x^i$,

one has $\delta(x_1x_2) = \delta(x_1)x_2^n + x_1^n\,\delta(x_2)$.

If

$$f = g[x_1]x_4 + g[x_2]x_3 + ax_3x_4 + \frac{1}{a}(g[x_1]g[x_2] - g[x_1x_2]),$$

where $g[x] \in F[x]$ and $a \in F^*$, then for $\delta = ad + p$, where $p[x] = g[x]$, one has $\delta(x_1x_2) = \delta(x_1)\,\delta(x_2)$. This completes the proof of theorem 1'.

If in addition the operation d: $R \to R$ is an additive map then the equality $d((x+y)z) = d(xz) + d(yz)$ draws in equality

$$f[x+y, z, d(x) + d(y), d(z)] = f[x, z, d(x), d(z)] + f[y, z, d(y), d(z)].$$

Therefore finding the solutions of the following system

$$\left| \begin{array}{l} f[x_1x_2, x_3, f[x_1, x_2, x_4, x_5], x_6] = f[x_1, x_2x_3, x_4, f[x_2, x_3, x_5, x_6]] \\ f[x_1, x_2, x_3, x_4] = f[x_2, x_1, x_4, x_3] \\ f[x_1 + x_2, x_3, x_4 + x_5, x_6] = f[x_1, x_3, x_4, x_6] + f[x_2, x_3, x_5, x_6] \end{array} \right. \tag{5}$$

of polynomial-functional equations is natural.

Theorem 2.

System (5) has only the following solutions in $F[x_1, x_2, x_3, x_4]$:

$f = ax_1x_2$, *where* $a \in F$

$f = x_1x_4 + x_2x_3 + ax_1x_2$, *where* $a \in F$

$f = ax_3x_4 + b(x_1x_4 + x_2x_3) + \dfrac{b(b-1)}{a}x_1x_2$, where $a \in F^*$, $b \in F$

Proof of Theorem 2. To prove this theorem the obtained solutions of system (1) can be used. But here we offer an easier way: Indeed, due to (2) the 3rd equality of system (5) can be written in the following equivalent form:

$$\begin{cases} f_0[x_1 + x_2, x_3] = f_0[x_1, x_3] \\ f_1[x_1 + x_2, x_3] = f_1[x_1, x_3] = f_1[x_2, x_3] \\ f_2[x_1 + x_2, x_3] = f_2[x_1, x_3] + f_2[x_2, x_3] \\ f_3[x_1 + x_2, x_3] = f_3[x_1, x_3] + f_3[x_2, x_3] \end{cases}$$

The first (second; third; fourth) equality of this system means that $f_0[x_1,x_2]= g_0[x_2]$ for some $g_0[x] \in F[x]$ (corresp. $f_1[x_1,x_2] = g_1[x_2]$ for some $g_1[x] \in F[x]$; $f_2[x_1,x_2] = x_1g_2[x_2]$ for some $g_2[x] \in F[x]$; $f_3[x_1,x_2] = x_1g_3[x_2]$ for some $g_3[x] \in F[x]$). In other words one has

$$f[x_1,x_2,x_3,x_4] = x_3x_4g_0[x_2] + x_3g_1[x_2] + x_4g_2[x_2] + x_1g_3[x_2].$$

Afterwards due to the 2nd equality of system (5) one has

$$\begin{cases} g_0[x_1] = g_0[x_2] \text{ i.e. } g_0[x] = a \in F \\ g_1[x_2] = x_2g_2[x_1] \\ x_1g_2[x_2] = g_1[x_1] \\ x_1g_3[x_2] = x_2g_3[x] \; i.e. \; g_3[x] = \text{ for some } c \in F \end{cases}$$

The 2nd and 3rd equalities of the last system imply $g^2[x]= b \in F$, $g_1[x] = bx$. Therefore

$$f[x_1,x_2,x_3,x_4] = ax_3x_4 + b(x_3x_2 + x_4x_2) + cx_1x_2$$

Now it is easy to see that for such $f[x_1,x_2,x_3,x_4]$ the 1st equality of system (5) is valid if and only if

$$f[x_1,x_2,x_3,x_4] = cx_1x_2 \text{ or } f[x_1,x_2,x_3,x_4] = x_1x_4 + x_2x_3 + cx_1x_2 \text{ or}$$

$$f[x_1,x_2,x_3,x_4] = ax_3x_4 + b(x_1x_4 + x_2x_3) + \frac{b(b-1)}{a}x_1x_2, \text{ where } a \in F^*$$

This completes the proof of theorem 2.

One can check by direct calculation or use Theorem 2 be sure that if f is a solution for system (5) then

$$\bar{f}[x_1, x_2, x_3, x_4] = af[x_1, x_2, a^{-1}(x_3 - cx_1), a^{1}(x_4 - cx_2)] + cx_1x_2$$

is also solution for system (5) for any $a \in F^*$ and $c \in F$. In other words the set of all solutions of system (5) is invariant with respect to the following action of the group $G_2 = F^* \rtimes F$ on it:

$$((a,c), f[x_1, x_2, x_3, x_4]) \mapsto af[x_1, x_2, a^{-1}(x_3 - cx_1), a^{-1}(x_4 - cx_2)] + cx_1 x_2.$$

It corresponds to the fact that if d: $R \to$ R is such an additive operation that $d(xy)$ is a polynomial in x, y, dx, dy over F and $a \in F^*$, $c \in F$ then $\delta(xy)$ is also a polynomial in x, y, δx, δy over F, where $\delta = ad + c$ i.e. $\delta(x) = ad(x) + cx$ by definition, moreover δ is also an additive operation.

The following result gives the "simplest" forms of $d(x_1x_2)$ with respect to that action.

Theorem 2·.

With an accuracy of the above action of group G_2 there are only the following equalities:

1) $d(x_1x_2) = o$, which corresponds to $f = 0$,
2) $d(x_1x_2) = d(x_1)x_2 + x_1d(x_2)$, which corresponds to $f = x_3x_2 + x_1x_4$,
3) $d(x_1x_2) = d(x_1)d(x_2)$, which corresponds to $f = x_3x_4$.

Proof of Theorem 2·. If $d(x_1,x_2) = cx_1x_2$ where $c \in F$ then for $S = d\text{-}c$, one has $\delta(x_1x_2) = 0$.

If $d(x_1x_2) = d(x_1)x_2 + x_1d(x_2) + cx_1x_2$, where $c \in F$ then for $\delta = d + c$ one has $\delta(x_1x_2) = \delta(x_1)x_2 + x_1\delta(x_2)$

If $f[x_1,x_2,x_3,x_4] = ax_3x_4 + b(x_1x_4 + x_2x_3) + \frac{b(b-1)}{a}x_1x_2$, where $a \in F^*$ then for $\delta = ad + b$, one has $\delta(x_1x_2) = \delta(x_1)\,\delta(x_2)$. This completes the proof of theorem 2·.

Theorem 2· can be considered as a confirmation of special positions of differential operators and homomorphisms in theory of commutative associative rings. Roughly speaking, Theorem 2· says that they exhaust all additive maps d:R $\to$ R for which $d(xy)$ is polynomial in $x, y, d(x), d(y)$.

Of course an analogical problem can be considered for other types of algebras, for example, associative algebras, Lie algebras or, in general, polynomial algebras (Procesi 1973). It would be interesting to investigate this problem for associative algebras.

REFERENCES

WAERDEN, B.L.VAN DER. 1991. *Algebra.* New York: Springer-Verlag.v.1.

PROCESI,CLAUDIO. 1973. *Rings with Polynomial Identities.* New York: Marcel Dekker,Inc.

Pertanika J. Sci. & Technol. **10(1)**: 119 - 134 (2002)
ISSN: 0128-7680

Critical Success Factors Influencing Adoption of Internet Technology by MSC & Non-MSC Companies

Lloyd Tam Yew King
Business Advanced Technology Centre, Universiti Teknologi Malaysia
E-mail: tyking@pc.jaring.my

Received: 30 March 2001

ABSTRAK

Internet dipercayai mengubah corak strategi perniagaan dibentuk dalam dunia korporat. Ia juga bukti sebagai alat strategik yang paling efektif dan berpengaruh dalam abad ini dengan menghubungkan ribuan syarikat dan jutaan orang setiap minit. Kertas kerja ini mengemukakan faktor-faktor utama yang digunakan oleh syarikat MSC dan bukan MSC dalam penerimaan teknologi Internet. Lima faktor organisasi dan dua faktor pasaran digunakan dalam kajian ini. Hasil kajian ini menyokong hipotesis bahawa di antara syarikat MSC dan bukan MSC, wujud perbezaan pendapat yang nyata terhadap enam faktor iaitu tanggapan faedah langsung, kesesuaian organisasi, kerumitan teknikal, tekanan rakan perniagaan, sokongan organisasi, dan tanggapan faedah tidak langsung. Kesemua faktor-faktor di atas selain daripada kerumitan teknikal menunjukkan bahawa syarikat bukan MSC adalah lebih dipengaruh oleh faktor–faktor ini daripada syarikat MSC. Satu lagi faktor iaitu tekanan persaingan didapati bukan faktor pengaruh.

ABSTRACT

The Internet is transforming the way in which business strategies are formulated in the corporate world. Indeed, it is proving to be the most effective and influential strategic tool of the century by connecting thousands of companies and millions of people every minute. This paper highlights the salient factors influencing the adoption of Internet technology by MSC and Non-MSC companies. Five organisational and two marketplace factors were used in this study. The final analysis confirmed that MSC and Non-MSC companies were influenced by six factors — perceived direct benefits, organisational compatibility, technical complexity, trading partner pressure, organisational support, and perceived direct benefits. All of the factors above except for technical complexity proved that Non-MSC companies were more affected by these factors than MSC companies. Another factor, competitive pressure, was not an influential factor.

Keywords: Electronic commerce, Internet technology, World Wide Web

INTRODUCTION

In Malaysia, the Internet is increasingly popular among local and multinational companies as a medium to advertise their business and enhance corporate performance on the electronic superhighway. The Malaysian government has provided several incentives to encourage the growth of multimedia companies For instance, the Multimedia Super Corridor (MSC) is a pilot project to harmonise Malaysia with the global forces shaping the information age. It is also a perfect environment for companies wanting to create, distribute and employ multimedia products and services. Malaysians cannot avoid the MSC and Internet technologies as both are already here. Companies have to prepare themselves to exploit the technologies of the Internet. In short, Internet technologies will become a way of life in Malaysia through the realisation of the MSC.

However, limited research had been done to identify the critical success factors (CSFs) use in adopting new innovations. In particular, information on salient factors that influence organisations' perception towards Internet technology adoption and its effect on both MSC and Non-MSC companies.

This paper is organised as follows: the next section presents the objective for this study, followed by a brief review of the literature. Subsequently, the hypotheses, methodology and results are presented. This is followed by an analysis of the hypotheses and discussion that relates the findings to the original theory.

OBJECTIVE OF STUDY

One of the most difficult challenges facing senior information systems (IS) managers is establishing a flexible IS infrastructure that will allow their organisations to successfully compete (Brancheau and Wetherbe 1996; Neiderman *et al.* 1991). Senior IS managers must work closely with functional business managers to establish an IS infrastructure that effectively supports existing systems, while remaining responsive to the constantly changing IS needs of the organisation (Stephens 1992). Establishing a stable IS infrastructure is not an easy task; numerous critical success factors, both internal and external to the firm, influence their decisions.

In recent years, the problem of establishing a stable IS environment is affected by a rapidly changing technical environment. The IS manager must decide whether to adopt emerging technologies based on both internal and external factors and on the firm's business objectives. Internal factors include top management support (Gordon and Gordon 1992), perceived benefits (Banerjee and Golhar 1994), technical compatibility (Iacovou *et al.* 1995), complexity (Rogers 1983), and organisational compatibility (Grover and Teng 1992). These impact both the type of technology adopted and the diffusion of the technology throughout the organisation. As firms attempt to leverage investment in information technology, the IS manager must determine whether emerging technologies support the firm's business plans and are compatible with the existing infrastructure. Several factors that are external to the organisation also impact the adoption decision. These include market competitiveness (Kunnathur *et al.* 1996) and pressure from other trading partners (Grover 1993). Thus, the objective of the study is to find answers to the following research questions.

RQ1: MSC companies that perceive greater benefits from adopting Internet will more likely adopt Internet technology than Non-MSC companies

RQ2: MSC companies that perceive Internet technology as compatible with their existing beliefs and work practices will more likely adopt Internet technology than Non-MSC companies.

RQ3: MSC companies that perceive Internet technology as compatible with their existing information systems environment will more likely adopt Internet technology than Non-MSC companies.

RQ4: MSC companies that perceive the adoption of Internet technology as a less complex process will more likely adopt Internet technology than Non-MSC companies

RQ5: MSC companies that receive pressure from their key trading partners to adopt Internet technology will more likely adopt Internet technology than Non-MSC companies.

RQ6: MSC companies with top management support for the adoption of Internet technology will more likely adopt Internet technology than Non-MSC companies.

RQ7: MSC companies that are in a highly competitive environment will more likely adopt Internet technology than Non-MSC companies.

LITERATURE REVIEW

On August 1, 1996, Prime Minister Datuk Seri Dr. Mahathir Mohamad announced the creation of a new urban zone designed specifically to enhance and develop a world-class multimedia industry and provide all necessary services to that industry. The MSC will be a test site that seeks to look into the possibility of integrating IT in everyday life, from business to government and to individuals. The MSC will incorporate not only infrastructure but also the business and legislative aspects of IT. It will not only benefit Malaysia but also the rest of the world (Computimes Series 1997-2000). This initiative is important for more reasons than one. From an economic point of view, MSC is important in helping Malaysia to achieve Vision 2020, which is the deadline set for the country to achieve fully developed nation status.

Presently, the country is undergoing a change from an industrial to information age and the MSC is providing a perfect environment for companies wanting to create, distribute and employ multimedia products and services. Multimedia Development Corporation (MDC) was set up to promote, implement, co-ordinate and manage the MSC (Computimes series 1997-2000).

In short, the MSC was introduced as a catalyst for the development of the latest information-based industries and is an area developed to encourage mutual enrichment of companies using modern technologies in a borderless world. The establishment of the MSC will enable Malaysia to leapfrog into the Information Age. MSC status companies are targeted to contribute about 50 percent or RM460 billion to the nation's gross domestic product (GDP) by 2020. To help meet this objective, MDC is trying to make it as easy as possible for MSC status companies to raise funds to support growth (i.e. fund projects, hire-experts, adopt hi-tech equipments) . One of the instruments established for this purpose was the Malaysian Exchange of Securities Dealing and Automation Quotation (Mesdaq) (Computimes series 1997-2000). The Malaysian government has also provided several incentives to encourage the growth of multimedia companies.

MSC-status companies incorporated in Malaysia can be wholly-owned by local and foreign legal entities. As of April, 2000, a total of 343 companies have been awarded multimedia status in software development (36 %), systems integration (17 %), content development (17%), telecommunication (11%) and other IT-related activities (19%), out of which 206 companies or 59 percent is local, 13 percent from Europe, 8 percent from United States and the rest from other countries. In the future, MSC hopes to house about 500 of the world class companies by 2020. It expects to have at least 50 world-class companies by 2003, increasing to 250 by 2010 (Computimes series 1997-2000).

The MSC is the beginning of a new era for Malaysia as it pave the way to a technology-based economy and a thinking society. The flagship applications of the MSC are actually Internet technology-based. As the dawning of the Internet technology era is taking place around the world, Malaysia is moving along with the MSC initiatives to attract the best-of-the-best technology, services and businesses around the world.

The realisations of electronic business will be one of the first things that everyone will encounter, even in Malaysia. This is the borderless market. Anyone can do business at anytime, anywhere. Of course, legal and business issues may abound but exploitation of the Internet technologies will overcome any challenges posed. Internet technology must also address the means of providing secure integrated, flexible, business-critical applications that create new value for businesses on the Internet. MSC and Non-MSC companies must be confident of these means in doing businesses with others as well. These new technologies and applications are now becoming very prevalent and they provide new way to make them available to employees, trading partners and customers.

Internet technologies will let Malaysian IT companies to automate their business processes from end-to-end, reducing costs and cycle time, and giving them tremendous boost in efficiency. It will also enhance the company image, communication and services by leveraging information. This is nor merely a Web home page or construction a browser. It is the integration of business processes into the Internet. Finally, the vast reach of the Internet gives them an opportunity to increase revenue by opening up near markets and providing them with new electronic channels. They can now extend the reach and range of physical boundaries and even right into the homes.

However, Tengku Datuk Dr. Mohd. Azzman Shariffadeen, National IT Council (NITC) permanent secretary advised companies planning to make investments into the MSC should not be taken by the hype created by the extensive press coverage on the subject, which have created a kind of gold rush among companies keen to get onto the bandwagon. "Companies must differentiate the hype from the facts before making decisions," he said, reminding that the MSC will not be a "golden mountain" unless the companies are able to utilise new technologies in smart ways (Computimes series 1997-2000).

With the emergence of interorganisational IS such as electronic data interchange (EDI) and electronic mail (e-mail) has allowed firms to adopt business strategies designed to leverage the speed with which they transmit critical business information (Johnston and Vitale 1988). By actively participating in integrated business strategies (e.g. Just-In-Time and Continuous Replenishment) in which all members of the value chain (e.g. suppliers, manufacturers, transporters and retailers) share time sensitive business information) electronically, firms have found they can reduce transaction and inventory costs and improve overall customer service (Srinivasan *et al.* 1994).

Over the past few years, e-commerce has attracted widespread interest from both functional and IS management. The steady growth in sophistication and functionality of advanced hardware, software, networking, and telecommunications technologies has provided top managers the opportunity to reassess the way they are currently doing business (Kalakota and Whinston 1997; Steinfield *et al.* 1996). By leveraging the speed and connectivity that are an inherent part of doing business electronically, firms have realised both operational and administrative benefits that can improve the organisation's competitive position (Benjamin *et al.* 1990; Kalakota and Whinston 1997).

The concept of using electronic networks to effectively communicate and share information both with trading partners and consumers is still in its infancy; therefore e-commerce is vaguely defined. However, one definition that seems to effectively capture its broad scope states that e-commerce is "a new way of conducting business characterised by companies and their customers performing electronic transactions through computer networks" (Cronin 1994).

Based on this definition, any electronic technology that supports the timely movement of critical business information from one party to another to facilitate a business transaction is electronic commerce. Thus, e-commerce includes electronic data interchange (EDI), smart card, symbol technology, bar coding, interenterprise messaging and file transfer and the World Wide Web (Pyle 1996). By exploring new ways to utilise these technologies to improve or enhance existing business processes, firms have begun to reap the benefits of doing business in an electronic environment.

Hypotheses

Innovation adoption research indicates that an organisation will only choose to adopt an innovation if it perceives that doing so will provide significantly greater benefits than existing technologies and processes (Rogers 1983). The organisation must perceive that the adoption of the innovation will either resolve existing operational problems or provide the firm with new business opportunities.

H1: The means for perceived benefits are greater for MSC status companies than Non-MSC status companies

Organisational compatibility is the extent to which a technology is consistent with the values, needs or experiences of the organisation (Rogers 1983). Process oriented compatibility is the extent to which an innovation is congruent with the exiting practices and processes of the firm (Tornatsky and Klein 1982).

H2: The means for organisational compatibility are greater for MSC status companies than Non-MSC status companies

If a new technology is incompatible with the firm's existing values, preferred work practices, or existing IS infrastructure, it is less likely to be adopted (Kwon and Zmud 1987). This is crucial point, because the adoption of Internet technologies often requires firms to modify existing business practices to realise benefits (Jarvenpaa and Ives 1996). There is generally a positive relationship between compatibility and adoption behaviours (Ettlie *et al.* 1984; Ettlie and Vallenga 1979).

H3: The means for technical compatibility are less for MSC status companies than Non-MSC status companies

Complexity is the degree of difficulty that users will experience when trying to understand or use an innovation or technology in the workplace (Kwon and Zmud 1987; Rogers 1983). The introduction of a new technology can be a complex and intimidating process for firm employees, particularly if the technology requires them to change their existing business practices or acquire new technical skills.

H4: The means for complexity are less for MSC status companies than Non-MSC status companies

The Internet facilitates the sharing of information between businesses. In order for an organisation to fully realise the benefits associated with the adoption of these technologies, it is essential that a significant number of other firms with which it shares information (trading partners) also adopt the technology (Stephens 1992). Firms that have fully adopted Internet technology will exert pressure on other trading partners to adopt (Davis 1995). Depending on that firm's power over trading partners and the extent of vertical dependence between firms in the value chain, organisations may be pressured to adopt a technology (Provan 1980).

H5: The means for trading partner pressure are greater for MSC status companies than Non-MSC status companies

Strong support of the top managers is vital to innovation adoption (Ettlie *et al.* 1984; Lederer and Mendelow 1988; Zmud 1984). Top management support goes beyond general approval for technology acquisition and includes a strong commitment from top management to support the technology at all levels of the organisation (Lederer and Mendelow 1988). Research indicates that securing top management support is a good predictor of level of success of a new information technology (Ives and Olson 1984).

H6: The means for top management support for the adoption of Internet technology are greater for MSC companies than Non-MSC companies

Because today's market place is increasingly competitive for many industries, firms are willing to explore the adoption of innovations in an attempt to gain a competitive advantage (Porter 1990). In an environment where the firm perceives a high level of competitive intensity and rivalry, the firm is more likely to allocate funds for the adoption of innovations; resulting in a greater level of overall innovation within the firm (Kimberly and Evanisko 1981; Reich and Benbasat 1990).

H7: The means for competitive pressure are greater for MSC status companies than Non-MSC status companies

METHODOLOGY

The research design for the study is exploratory in nature. Exploratory research was designed to provide a summary of some aspects of the environment when the hypotheses were tentative and speculative in nature (Aaker and Day 1990).

Data Collection Method

The data was secured by means of questionnaires, distributed to both MSC and Non-MSC companies that were planning to adopt, currently adopting and those that had already adopted Internet technology. Specifically, eligible respondents consisted of top IT executives who are responsible for managing the assessment and adoption of innovative information systems technologies.

According to the "*Computer Era*" directory, the population number of public and private IT organisations in Malaysia is about 1,976. The target population for this study was the organisations in the Malaysian IT industry. Selangor and

Kuala Lumpur, being the popular places among IT companies (73.5%), were selected as the location of study. The final sample size consisted of 306 respondents selected from records listed in the directory via a simple random sampling process. A set of self-administered questionnaire was handed to a potential respondent that satisfied the survey criteria and returned once the respondent finished answering it. The questionnaires were kept as simple, short and self-explanatory as possible.

Out of the total number of distributed, 250 usable questionnaires were obtained for analysis. The remaining number was deleted because of incomplete data (6), non-respondents (25), and those that did not adopt the technology (25). Non-response bias between respondents and non-respondents was also tested using chi-square test and the result showed there no significant difference at _ = 0.05 significance level for any of the respondent's demographic variables.

Data Analysis Techniques

Data were analysed using descriptive statistics, factor analysis, Cronbach's coefficient alpha, and MANOVA. To analyse the respondent's background, descriptive analysis and common measures such as *total, mean, frequencies* and *percentage* were utilised. The demographic information included: (I) job title, (II) the respondent's level of experience, (III) MSC status, (IV) total number of employees, (V) total number of IS employees, (VI) percent of firm's budget dedicated to IS, (VII) use of other Internet technologies, and (VIII) organisation's age.

Next, factor analysis was used to assess unidimensionality and Cronbach's coefficient alpha was used to assess internal consistency. The researcher then interpreted the output and related the findings to the hypotheses. The study consists of seven research hypotheses that were tested using MANOVA technique to determine whether the research hypotheses were supported by the data collected.

Respondent Profile

A series of eight questions were used to obtain demographics information on both the respondent and the respondent's organisation (Table 1).

FINDINGS

An exploratory factor analysis was used to help assess the unidimensionality of the multi-item scales. The unidimensionality of a set of items used to measure a given construct is necessary, but nor sufficient, condition for construct validity. Construct validity was also assessed by examining the internal consistency, and convergent and discriminant validity of each construct.

A principle components factor analysis using a Varimax rotation was performed using the twenty-nine items proposed to measure the following seven constructs: top management support, organisational compatibility, technical compatibility, complexity, competitive pressure, trading partner pressure, and

TABLE 1
Respondent and organisational profile

Category	Frequency (N)	Valid Percentage (%)
Job title		
IT managers	51	20.4
Chief Information Officers	47	18.8
Vice-Presidents	29	11.6
Director of IS	22	8.8
Others	101	40.4
Internet technology experience level		
Very experience	47	18.8
Somewhat experience	72	28.8
Experience	63	25.2
Limited experience	58	23.2
Not experience	10	4.0
MSC status		
Non-MSC companies	126	50.4
MSC companies	124	49.6
Total employees		
Less Than 10 (Micro)	68	27.2
More Than 10 Up To 100 (Small)	97	38.8
More Than 100 Up To 500 (Medium)	46	18.4
More Than 500 (Large)	39	15.6
Total IS employees		
Less than 3	73	29.4
More than 3 up to 10	91	36.7
More than 10 up to 50	29	11.7
More than 50	55	22.2
Missing	2	
Annual IS budget		
Less than 1%	82	36.4
More than 1 % up to 5%	113	50.2
More than 5% up to 10%	20	9.0
More than 10%	10	4.4
Missing	25	
Internet technologies		
Electronic Mail (E-mail)	241	91.6
Web-site	200	80.0
Intranet	199	79.6
Extranet	168	67.2
Electronic Data Interchange (EDI)	57	22.8
Electronic Commerce (E-commerce)	56	22.4
Electronic Fund Transfer (EFT)	27	10.8
Years company in operation		
Less than 1	38	15.2
1 to less than 5	136	54.4
5 to less than 10	55	22.0
10 years to less than 20	17	6.8
More than 20	4	1.6

perceived benefits. The criteria used to determine the number of factors to extract was an eigenvalue that was greater than equal to one (Jeller and Carmines 1980). The results indicated that seven factors had eigenvalues exceeding 1.00 (Table 2). Thus, seven factors were extracted during this analysis.

Dimensionality of each of the factors was assessed by examining the factor loadings. Items with factor loadings of greater that 0.5 on the factor with which they are hypothesized to load were considered adequate indicators of that factor (Hair *et al.* 1995). However, items with factor loadings of at least 0.3 on other factors were examined to see it they measured an additional factor.

Based on the rotated results, a new construct called organizational support was developed which encompassed both the three items in the hypothesized construct top management support and two items from organisational compatibility. Another new construct called technical complexity was developed which encompassed both the items in the hypothesized construct technical compatibility and complexity. As for the eight items hypothesized to measure the construct perceived benefits were loaded on two factors, perceived direct benefits and perceived indirect benefits.

Reliability has been defined as the "degree to which measures are free from error and therefore yield consistent results" (Peter 1981). One aspect of reliability is internal consistency which is an indicator of the level of homogeneity of a measuring scale (Cronbach 1951). One criterion that has been widely used to assess the reliability of a multi-item measurement scale is Cronbach's (Cronbach 1951) coefficient alpha. Based on the reliability analysis result, six of the seven constructs had coefficient alpha values exceeding 0.7 (see Table 2). Only the construct competitive pressure had coefficient alpha of 0.54. Nunnally (Nunnally n.d) suggested that a set of items with a coefficient alpha greater than 0.7 is considered internally consistent. Because this construct had a coefficient alpha less then the recommended level of 0.7, its internal consistency was weak. Therefore, it was not used in subsequent analysis.

Test of Hypotheses

MANOVA was used to examine the relationship between companies' MSC status and the final six constructs hypothesized to impact Internet technology adoption within organisations. Based on the findings, it can be seen that there is significant differences between companies' MSC status and factors for adopting Internet technology. The Wilk's Lambda, (F-value = 44.169), and the level of significance, (p-value = 0.0001 < 0.05), indicate that the means for the MSC and Non-MSC companies contained significant differences at the a = 0.05 level (Table 3).

Table 4 shows a comparison of adoption factors between MSC and non-MSC companies by using their respective mean values. The mean values for the six Internet technology adoption factors are ranked in a definite order and followed the sequence of importance, beginning with most important to least

TABLE 2
Confirmation of the 7 factors

Measurement Variable and Dimension	Factor Loading
Factor 1: Organisational Support (Reliability α = 0.8900)	
Organisation Values and Beliefs	0.838
Top Management Communicate	0.799
Top Management Interest	0.797
Favourable Attitude	0.797
Top Management Importance	0.788
Factor 2: Trading Partner Pressure (Reliability _ = 0.9123)	
Trading Partner Business Needs	0.890
Adversely Impart Trading Partner Relations	0.849
Trading Partner Strategies	0.826
Trading Partner Recommendation	0.815
Factor 3: Perceived Direct Benefits (Reliability _ = 0.8877)	
Reduce Transaction Costs	0.899
Improve Overall Productivity	0.865
Improve Cash Flow	0.822
Improve Operational Efficiency	0.759
Factor 4: Technical Complexity (Reliability _ = 0.7953)	
Decrease Productivity-Time To Learn	0.818
Complex To Use	0.805
Disrupt Work Environment	0.783
Complex To Develop	0.680
Factor 5: Perceived In-direct Benefits (Reliability _ = 0.9173)	
Improve Existing Customer Relations	0.865
Reach New Customers	0.858
Increase Ability To Compete	0.837
Factor 6: Organisational Compatibility (Reliability _ = 0.8444)	
Computerised Data Resources	0.856
Organisational Experience	0.820
Communications Infrastructure	0.772
Factor 7: Competitive Pressure (Reliability _ = 0.5438)	
Customers Can Switch Easily	0.751
Intense Competitive Rivalry	0.726
Monitor Competitors Action	0.636

TABLE 3
MANOVA statistic results

Statistics	F-value	D.F.	p-value
Wilks' Lambda	44.169	7	0.0001
Pillai's Trace	44.169	7	0.0001
Hotelling's Trace	44.169	7	0.0001
Roy's Largest Root	44.169	7	0.0001

important adoption factors for MSC and Non-MSC companies. The adoption factors ranking identified for MSC organisations are similar to those for Non-MSC companies, with exception of organisational compatibility and perceived in direct benefits. Perceived in-direct benefits factor was second in importance for MSC companies and was third in importance for Non-MSC companies. The organisational compatibility factor was second in importance for Non-MSC companies and was third in importance for MSC companies. Other factors were considered to be similar in importance by MSC and Non-MSC companies.

TABLE 4
Comparison between MSC and Non-MSC companies

MSC companies	Means	Non-MSC companies	Means
1. Organisational Support	3.5258	Organisational Support	4.1778
2. Perceived In-Direct Benefits	3.0914	Organisational Compatibility	3.9339
3. Organisational Compatibility	2.8978	Perceived In-Direct Benefits	3.8624
4. Perceived Direct Benefits	2.8669	Perceived Direct Benefits	3.4583
5. Trading Partners Pressure	2.6714	Trading Partners Pressure	3.3710
6. Technical Complexity	2.4738	Technical Complexity	2.1667

NOTE: The scale used for all constructs ranged from 1 (strongly disagree) to 5 (strongly agree)

A further analysis between the companies' MSC status and each of the factors for adopting Internet technology shows that the F-values obtained from the univariate analysis of variance for each construct were significant (p-value > 0.05) for hypotheses, H1 (perceived direct benefits), H2 (organisational compatibility), H3 (technical complexity), H4 (trading partner pressure), H5 (organisational support), and H6 (perceived in direct benefits). Therefore, the null hypothesis (H_0) was rejected and the alternative hypothesis (H_1) was accepted. However, the univariate F-values for hypotheses, and indicated that the means for those constructs were significantly greater for non-MSC than for MSC companies. This proves that the expected results were not supported (Table 5). Hence, the mean-values were greater for non-MSC companies than for MSC companies.

CONCLUSION

The main objective of this study was to determine the CSFs that influences MSC and Non-MSC companies to adopt Internet technology. This study was carried out using a questionnaire survey on 306 randomly selected IT organisations located around Selangor and KL. Appropriate measure of control and precautions were taken during each section to produce more reliable and meaningful views and opinions. Through a pilot test, the reliability and consistency of the survey instrument were examined. Factor analysis was used to group and minimise 29 variables into 7 constructs for easier management. The

TABLE 5
Summary of research findings by MSC status

Constructs	F-value	p-value	Expected Results	Actual Results
Perceived Direct Benefits	42.260	0.000	Greater Means for MSC Companies than Non-MSC Companies	Not supported
Organisational Compatibility	168.528	0.000	Greater Means for MSC Companies than Non-MSC Companies	Not supported
Technical Complexity	11.852	0.001	Lower Means for MSC Companies than Non-MSC Companies	Not supported
Trading Partner Pressure	50.788	0.000	Greater Means for MSC Companies than Non-MSC Companies	Not supported
Organisational Support	58.335	0.000	Greater Means for MSC Companies than Non-MSC Companies	Not supported
Perceived Indirect Benefits	60.126	0.000	Greater Means for MSC Companies than Non-MSC Companies	Not supported
Competitive Pressure (*)			Greater Means for MSC Companies than Non-MSC Companies	Not supported

NOTE: * not included in MANOVA because the factor was not internally consistent

correlation of the constructs was also investigated. As a result, the competitive pressure construct was found to be not reliable; thus, it was dropped from further analysis.

The overall result shows that there was significant difference between MSC and Non-MSC companies and six of the seven constructs (i.e. organisational compatibility, technical complexity, organisational support, trading partner pressure, perceived direct benefits, perceived in-direct benefits). In short, both MSC and Non-MSC companies were influenced by these factors to adopt Internet technology.

However, the mean-values were significantly greater for Non-MSC and MSC companies, which means that the expected results were wrong and that Non-MSC companies were more concerned with the above factors than the MSC companies and vice-versa for the technical complexity construct. According to theory, Non-MSC companies are more likely to adopt Internet technology if it is compatible with their existing beliefs and work cultures than MSC companies because most of the MSC companies are small and fairly new; thus, they should face less problems to incorporate new technologies into their corporate cultures. As for the technical complexity issue, Non-MSC companies were less concerned about it than MSC companies as they have more financial resources to spend on hi-tech equipments or/and hire external consultants to solve any problems they might face. In addition, bigger Non-MSC companies have more IS staff per total company employees than smaller MSC companies. Non-MSC companies also have to communicate their intention to more internal and external parties (i.e. employees, trading partners, and customers) to gain support from them, otherwise none of them would know about it or use it. Since Non-MSC organisations have to deal with more trading partners, they tend to receive more pressure from them than MSC organisations. Finally, Non-MSC companies are more particular about the direct and indirect benefits from adopting Internet technology because any decision they make will affect their operations, productivity, competitiveness, customer relations, cash flow, and etc.

ACKNOWLEDGEMENT

The author would like to thank Dr. Nooh Abu. Bakar, professor of Business Advance Technology Centre, Universiti Technologi Malaysia for his guidance and encouragement in preparing this paper. Special thanks also go to Dr. Izhar Che Mee, Chief Executive Officer of I-Commerce Sdn Bhd. and General Manager of MRCB Sdn Bhd. for his invaluable support.

REFERENCES

AAKER, D.A. and G.S DAY. 1990. *Marketing Research.* 4th ed. John Wiley & Sons.

BANERJEE, S. and D. Y. GOLHAR. 1994. Electronic data interchange: characteristics of uses and nonusers. *Information and Management.* p. 65 – 74.

BENJAMIN, R. I., D. W. DELONG and M. S. MORTON. 1990. Electronic Data Interchange: How Much Competitive Advantage. Long Range Planning. p. 29 – 40.

BRANCHEAU, J. C. and J. C. WETHERBE. 1996. Key issues on information systems management: 1994-1995 SIM Delphi results. *MIS Quarterly.* p. 225 – 242.

COMPUTIMES SERIES 1997-2000. News Strait Times Press.

CRONBACH, L. J. 1951. Coefficient alpha and the structure of tests. *Psychometrika* September: 297 – 334.

CRONIN, M. J. 1994. *Doing Business on the Internet: How the Electronic Highway is Transforming America Companies.* New York: Van Nostrand Reinhold.

DAVIS, D. 1995. Evolution of EDI. *Manufacturing Systems* August: 22 – 28.

ETTLIE, J. E., W. P. BRIDGES and R. D. O'OKEEFE. 1984. Organizational strategy and structural differences for radical versus incremental innovation. *Managing Science.* p. 682 – 695.

ETTLIE, J. E. and D. B. VALLENGA. 1979. The adoption time period of some transportation innovation. *Management Science.* p. 429 – 443.

GROVER, V. 1993. Empirically derived model for the adoption of customer based inter-organisational systems. *Decision Science.* p. 603 – 640.

GROVER, V. and J. T. C. TENG. 1992. An examination of DBMS adoption and success in America organisations. *Information & Management.* p. 239 – 248.

GORDON, S. R. and J. R. GORDON. 1992. Organisational hurdles to distributed database management systems (DDBMS) adoption. *Information & Management.* p. 333 – 345.

HAIR, J. F., JR. R. E. ANDERSON, R. L. TATHAM and W. C. BLACK. 1995. *Multivariate Data Analysis.* Edgewood Cliffs, NJ: Prentice-Hall.

IACOVOU, C. L., I. BENBASAT and A. S. DEXTER. 1995. Electronic data interchange and small organisations: adoption and impact of technology. MIS *Quarterly.* p. 465 – 485.

IVES, B. and M. H. OLSON. 1984. User involvement and MIS success: a review of research. *Management Science* (**30.5**): 586 – 603.

JARVENPAA, S. L. and B. IVES. 1996. Introducing transformational information technologies: the case of the world wide web technology. *International Journal of Electronic Commerce.* p. 95 – 126.

JOHNSTON, H. R. and M. R. VITALE. 1988. Creating Competitive advantage with inter-organisational information systems. *MIS Quarterly.* p. 153 – 165.

KALAKOTA, R. and A. B. WHINSTON. 1997. *Electronic Commerce: A Manager's Guide.* Reading MA: Addison Westley.

KIMBERLY, J. R. and M. J. EVANISKO. 1981. Organisational innovation: the influence of individual, organisational and contextual factors on hospital adoption of chronological and administrative innovations. *Academy of Management Journal.* p. 689 – 713.

KUNNATHUR, A. S., M. U. AHMED and R. J. S. CHARLES. 1996. Expert systems adoption: analytical study of management issues & concerns. *Information & Management.* p. 15–25.

KWON, T. H. and R. ZMUD. 1987. Unifying the Fragmented Models of Information Systems Implementation. In *Critical Issues in Information Systems Research,* ed. R.J. Boland and R.A. Hirschheim, p. 227 – 252. New York: John Wiley Publishing.

LEDERER, A. L. and A. L. MENDELOW. 1988. Convincing top management of the strategic potential of information systems. *MIS Quarterly.* p. 526 – 536.

MUKHOPADHYAY, T., S. KEKRE and S. KALATHUR. 1995. Business value of information technology: a study of electronic data interchange. *MIS Quarterly.* p. 137 – 165.

NEIDERMAN, F., J. C BRANCHEAU and J. C. WETHERBE. 1991. Information systems management issues in the 1990s. *MIS Quarterly.* p. 474 – 499.

NUNNALLY, J. C. *Psychometric Theory.* New York: McGraw-Hill.

PETER, J. P. 1981. Construct validity: a review of basic and recent marketing practices. *Journal of Marketing Research.* p. 133 – 145.

PROVAN, K. G. 1980. Recognising, measuring and interpreting the potential/enacted power distinction in organizational research. *Academy of Management Review.* p. 549 – 559.

PORTER, M. E. 1990. *The Competitive Advantage of Nations.* New York: The Free Press.

PYLE, R. 1996. Electronic Commerce and the Internet, Communications of the ACM. p. 22 – 23.

REICH, B. H. and I. BENBASAT. 1990. An empirical investigation of the factors influencing the success of customer-oriented strategies systems. *Information Systems Research.* p. 325 – 347.

ROGERS, E. M. 1983. *Diffusion of Innovations.* New York: Free Press.

SRINIVASAN, K., S. KEKRE and T. MUKHOPADHYAY. 1994. Impact of electronic data interchange on JIT shipments. *Management Science.* p. 1291 – 1304.

STEINFIELD, C., R. KRAUT and A. PLUMMER. 1996. The impact of inter-organisational networks on buyers-seller relationship. *Journal of Computer Mediated Communication.* p. 1 – 19.

STEPHENS, C. S. 1992. Executive or functional manager? The nature of the CIO's job. *MIS Quarterly.* 449 – 467.

SWATMAN, P. M. and P. SWATMAN. 1992. EDI system integration: a definition and literature survey. *The Information Society.* p. 169 – 205.

TORNATSKY, L. G. and K. J. KLEIN. 1982. Innovations characteristics and innovation adoption implementation: a met analysis of findings. *IEEE Transactions on Engineering Management.* p. 28 – 45.

ZELLER, R. A. and E. G. CARMINES. 1980. *Measurement in Social Sciences: The Link Between Theory and Data.* New York: Cambridge University Press.

ZMUD, R. 1984. An examination of push-pull. Theory applied to process innovation in knowledge work. *Management Science.* p. 727 – 738.

Pertanika J. Sci. & Technol. **10(1)**: 135 - 143 (2002)
ISSN: 0128-7680

Studies of Sn Substitution on Ca and Cu Sites of Bi-Sr-Ca-Cu-O Superconducting System

H. Azhan[1], S. A. Halim[2], S. B. Mohamed[2], K. Azman[1], S. Y. S. Yusainee[1] & H. A. A. Sidek[2]
[1]*Materials Science Research Group,*
Universiti Teknologi MARA,
Kampus Jengka, 26400 Bandar Jengka,
Pahang Darul Makmur
[2]*Superconductors and Thin Film Laboratory,*
Physics Department,
Universiti Putra Malaysia,
43400 Serdang, Selangor, Malaysia

Received: 16 November 2000

ABSTRAK

Kesan penggantian Sn pada tapak Ca dan Cu secara serentak atau berasingan dalam sistem Bi-Sr-Ca-Cu-O telah dikaji dengan menggunakan kaedah pembelauan sinar-x dan teknik pengukuran rintangan masing-masingnya untuk penentuan struktur dan suhu genting, T_C. Secara amnya, semua sampel menunjukkan sifat normal logam pada suhu melebihi suhu T_{Conset}. Nilai $T_{C(R=0)}$ didapati menurun terhadap kepekatan Sn. Walau bagaimanapun, nilai $T_{c(R=0)}$ untuk sampel x=0.02 yang didop dengan Sn secara serentak pada tapak Ca dan Cu diperolehi pada 104 K. Suhu ini meningkat sebanyak 4 K jika dibandingkan dengan suhu $T_{C(R=0)}$ sampel tulen. Sampel yang didop dengan Sn, untuk kepekatan x=0.20, pada tapak Ca atau kedua-dua tapak Ca dan Cu menunjukkan dominasi fasa-fasa 2212 dan 2201. Dengan yang demikian ubahsuaian terhadap persekitaran Ca membantu pembentukan fasa T_C rendah. Pemerhatian ini disokong oleh maklumat yang diperolehi melalui corak pembelauan sinar-x. Puncak-puncak baru yang tidak dikenal pasti (bendasing) dan puncak-puncak yang merujuk kepada fasa rendah 2201 muncul pada sampel dengan kepekatan Sn melebihi x=0.15. Tiada puncak yang dipunyai oleh SnO_2 dikesan dan ini bermaksud bahawa Sn mungkin telah bertapak ke dalam struktur kristal sistem BSCCO ataupun membentuk bendasing.

ABSTRACT

The influence of Sn substitution on Ca and Cu sites in Bi-Sr-Ca-Cu-O superconductor system simultaneously or separately have been studied using x-ray diffraction (XRD) method and resistance measurement technique for the structural identification and determination of critical temperature, T_C respectively. Generally, all samples displayed a normal metallic behavior above T_{Conset}. The values of $T_{C(R=0)}$ decreased towards Sn concentration. However, the $T_{C(R=0)}$ value for x=0.02 sample doped simultaneously in Ca and Cu sites was observed at 104 K. The critical temperature increased by 4 K compared to that of the pure sample. Sample doped with Sn, for concentration of x=0.20, at Ca site or at both Ca and Cu sites show the dominance of the 2212 and 2201 phases. Hence, altering the Ca environment favours the formation of the low T_C phases. This observation was also supported by the information obtained from the XRD patterns. New unidentified peaks (probably impurities) and low phase peaks corresponding to 2201 phase existed for samples with Sn concentration above x=0.15. No peaks belonging to SnO_2 were detected implying that Sn probably has been incorporated into the crystalline structures of the BSCCO system or formed as impurities.

Keywords: Superconductor, BSCCO, doping

INTRODUCTION

There are three stable phases in BSCCO system represented by a general formula, $Bi_2Sr_2Ca_{n-1}Cu_nO_y$. They were commonly referred to as 2201 phase (n=1, T_C ~ 10 K), 2212 phase (n=2, T_C ~ 80 K) and 2223 phase (n=3, T_C ~ 110 K) (Seshu Bai and Ravi 1991). Hence the formation of phases was due to the increasing content of Ca and Cu. In 2223 phase, partial substitution of lead in Bi-site was used to stabilize the formation of the phase and hence increased the volume fraction of 2223 phase and raised the T_C up to 110 K (Pisass *et al.* 1990). The effect of Ba doping at Ca site, both elements belonging to the same alkaline earth-group but having different ionic radii, in $Bi_2Pb_{0.6}Sr_2Ca_{2-x}Ba_xCu_3O_\delta$ suppressed the $T_{C(R=0)}$ but gradually decreased its value from 104K to 88K for x=0 to x=0.1 respectively (Halim *et al.* 1999). A partial substitution of V in Cu site, reduced superconducting properties in $(Bi,Pb)_2Sr_2Ca_2Cu_{3-x}V_xO_y$ by destroying the phase coherent (Nkum 1998).

It was believed that lead existed as Pb^{2+} and replaced Bi^{3+} in the structure and improves the superconducting properties for low concentration of lead substituted in Bi site of BSCCO superconducting system. At high concentration, lead also existed as Pb^{4+} and replaced other sites. The addition of Sn, which exists in two oxidation states, Sn^{2+} and Sn^{4+}, on Bi sites has reduced the volume of unit cell and reduced the formation of 2212 phase (Jha and Mendiratta 1996). Since tin have the similar electronic structure to that of lead, the study will be done to investigate the influence of a partial substitution of tin on Ca and Cu separately and both sites simultaneously on the superconducting properties.

EXPERIMENTAL PROCEDURE

The Sn-doped samples were prepared from Bi_2O_3, PbO, $SrCO_3$, CaO, SnO_2 and CuO powders (each at least 99.9 % purity) in the correct stoichiometric amount to produce samples base on these formula: $Bi_{1.6}Pb_{0.4}Sr_2(Ca_{1-x}Sn_x)Cu_3O_\delta$, $Bi_{1.6}Pb_{0.4}Sr_2Ca2(Cu_{1-x}Sn_x)_3O_\delta$ and $Bi_{1.6}Pb_{0.4}Sr_2(Ca_{2-x}Sn_x)_2(Cu_{1-x}Sn_x)_3O_\delta$. The concentration of Sn doping were in the range of x=0.02, 0.05, 0.10, 0.15, and 0.20. These powders were milled together with an absolute ethanol in alumina pot for 24 hours and dried out in oven at 120°C for 6 hours. These samples were calcined at 800°C for 24 hours. Further calcination was done at 830°C for 24 hours after grinding the powders using mortar and pestle. Finally, the powders of this nominal composition were pressed into disk shape with the diameter = 1.2 cm and the thickness = 2 mm before firing for 150 hours at 850°C in a box furnace as a final sintering process.

The x-ray diffraction patterns were checked using Cu-Ka radiation in the range 2° up to 70° of 2θ angles with an angular step of 0.02°. The critical temperature, T_C was determined from electrical resistance measurements. The standard four-point probe technique was used for the electrical resistance measurement. The system fitted with a close cycle helium cooling system was fully computerised. Silver paint was used as electrical contacts between the samples and the probes.

RESULTS AND DISCUSSION

Normalised resistance at room temperature as a function of temperature for the samples doped with Sn in Ca site was showed in *Fig. 1.* All samples displayed a normal metallic behaviour above the T_{Conset} (transition temperature from metallic to superconductor). The $T_{C(R=0)}$ (zero resistance temperature) and T_{Conset} for the pure sample were obtained at 100 K and 110 K respectively. The widening of ΔT=10 K in this sample could be due to the existence of impurities. Formation of the low T_C phase favoured as the concentration of Sn increased from x=0.02 and x=0.05 as seen by the two-step features, thus reducing the $T_{C(R=0)}$. However, the two-step features were not observed for other concentration. The suppression of T_C values for the partial substitution of Sn in Ca site indicated that the strength of coupling between the grains of mixed phases has been decreased and hence, increased the weak-link (Nkum and Datars 1995).

The effect of Sn doping at Cu site on the temperature dependence of the resistance was displayed in *Fig. 2.* All samples, except sample with x=0.3, showed a normal metallic behaviour above the T_{Conset}. As observed for the curve belonging to x=0.02 sample, the resistive tail persists beyond $T_{C(R=0)}$ of the pure

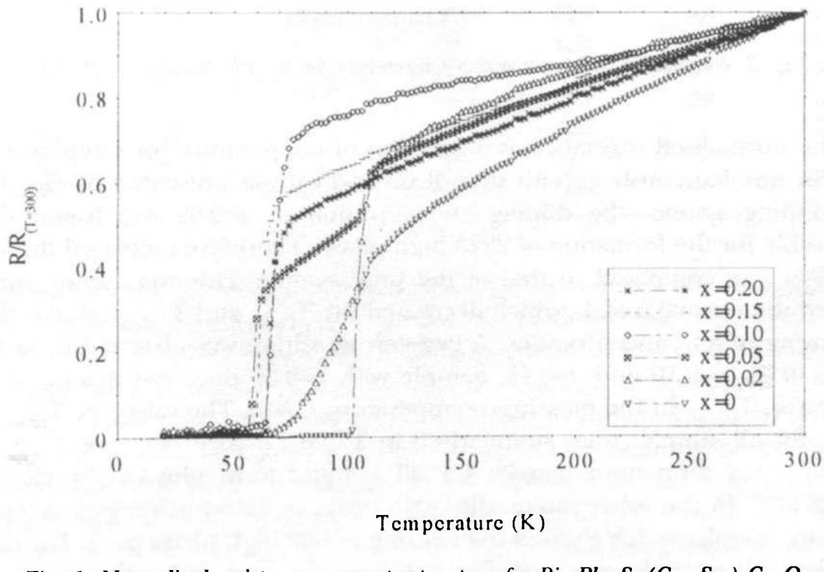


Fig. 1. Normalised resistance versus temperature for $Bi_{1.6}Pb_{0.4}Sr_2(Ca_{1-x}Sn_x)_2Cu_3O_\delta$

sample making ΔT very broad. This was due to the poor grain connectivity and the presence of low T_C phase. A two-step feature as clearly observed in all the samples except for the pure and sample with x=0.15. For samples with x=0.10 and x=0.15 the curves showed the improvement in $T_{C(R=0)}$ as compared to the samples doped in Ca site. It was a remarkable behaviour for the samples with x=0.20 and x=0.25 that showed a broadly improvement on the $T_{C(R=0)}$ and did not follow the trends of other samples that decreased the $T_{C(R=0)}$ with the increasing of Sn concentration. It was believed that the presence of resistive tail would be reduced after the heat treatment process implemented on these samples.

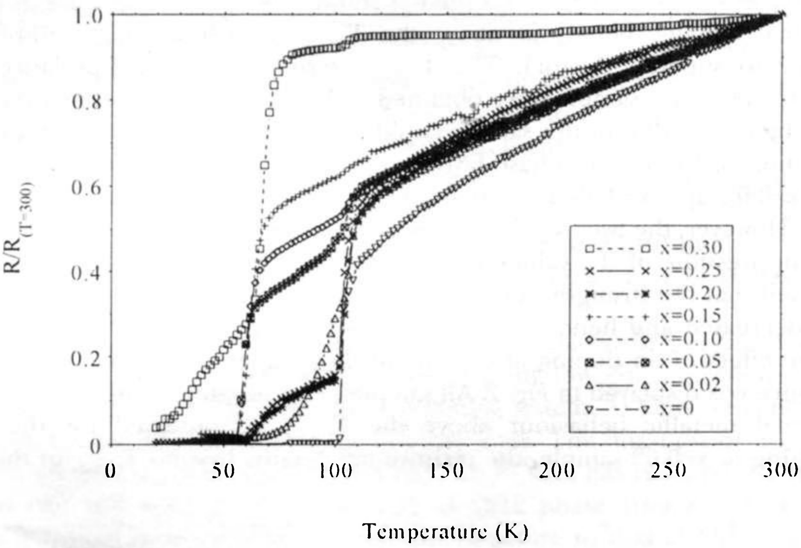


Fig. 2. Normalised resistance versus temperature for $Bi_{1.6}Pb_{0.4}Sr_2Ca_2(Cu_{1-x}Sn_x)_3O_\delta$

The normalised resistance as a function of temperature for samples doped with Sn simultaneously at both sites of Ca and Cu was presented by *Fig. 3.* For this doping system, the doping concentration at x=0.02 was found to be favourable for the formation of 2223 high phase. Therefore increased the T_{Conset} and $T_{C(R=0)}$ as compared to that of the pure sample. However, other samples showed the normal trends, which decreased the T_{Conset} and $T_{C(R=0)}$ value with the increasing of Sn concentrations. A two-step transition was observed in samples with x=0.05, x=0.10 and x=0.15. Sample with x=0.20 does not achieve a zero resistance, $T_{C(R=0)}$ in the measured temperature range. The values of T_{Conset} and $T_{C(R=0)}$ for all samples were summarised in Table 1.

The x-ray diffraction pattern for all samples were plot in the range of $20° < 2\theta < 40°$. In the other range, the main peaks remain unchanged except for the pure sample, which showed the existing of 002 high phase peak. For doped samples, this peak disappeared. *Fig. 4* showed the x-ray diffraction pattern for

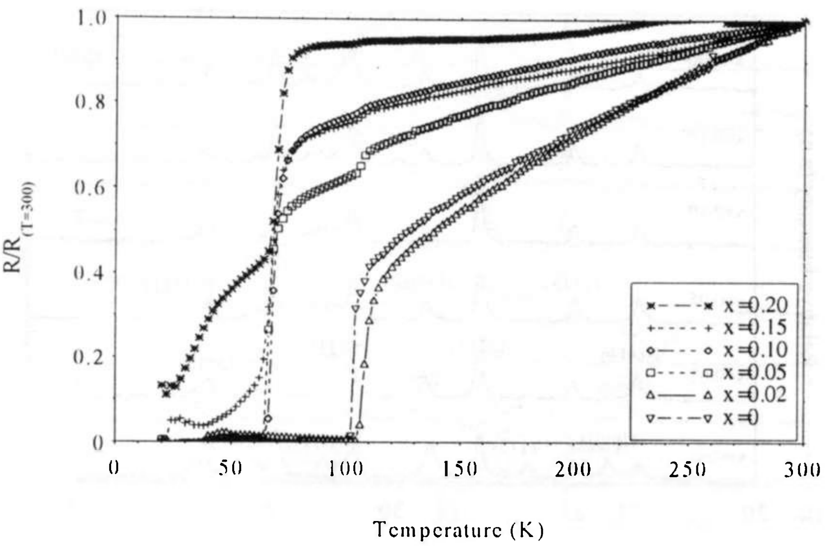


Fig. 3. Normalised resistance versus temperature for $Bi_{1.6}Pb_{0.4}Sr_2(Ca_{1-x}Sn_x)2(Cu_{1-x}Sn_x)_3O_8$

TABLE 1

T_{Conset} and $T_{C(R=0)}$ for all samples with a different concentration of Sn doped in Ca site, Cu site and both sites simultaneously

Sample Concen.	Sn doped in Ca site		Sn doped in Cu site		Sn doped in both sites	
	$T_{C(R=0)}$	T_{Conset}	$T_{C(R=0)}$	T_{Conset}	$T_{C(R=0)}$	T_{Conset}
0.00	100	110	100	110	100	110
0.02	60	112	66	114	104	114
0.05	56	110	54	110	64	110
0.10	64	76	56	108	64	80
0.15	58	70	54	108	58	70
0.20	62	72	54	110	-	78
0.25	-	-	58	112	-	-
0.30	-	-	23	106	-	-

the samples doped with Sn in Ca site. It can be observed clearly that 2223 high phase peaks dominated the pure sample. In Sn doped samples, some of the high peaks such as H(115) and H(00<u>10</u>) disappeared. Some of the peaks in the Sn doped samples were shifted such as peaks H(00<u>12</u>) as compared to that of the pure sample. One unidentified peak existed in the Sn doped samples and was marked as (*). This peak probably belongs to the impurity phase and it is believed that this peak correspond to the non-superconducting phase. The H(00<u>16</u>) and LL(115) appeared in samples doped with x=0.15 and x=0.20. However, no peaks belongs to SnO_2 were detected implying that this oxide was incorporated into the crystalline structure (Nkum and Datars 1995; Melghit *et*

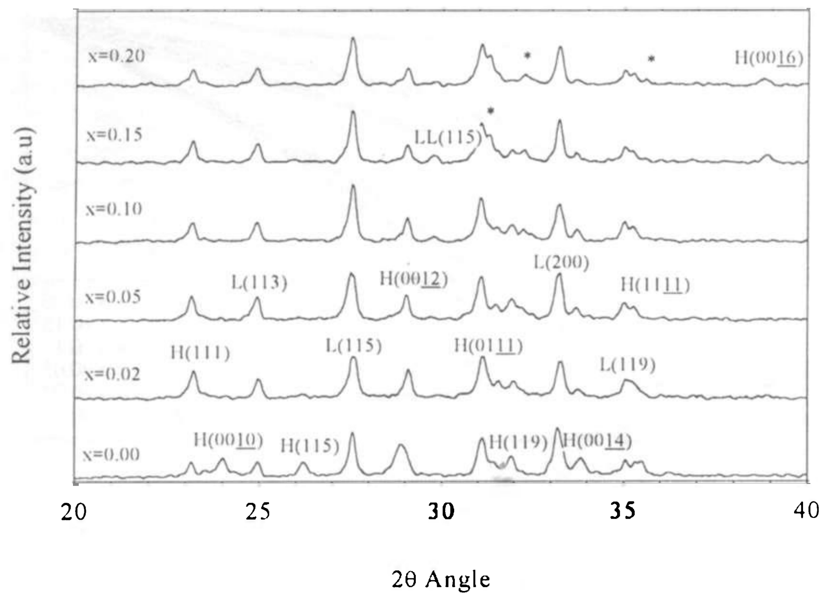


*Fig. 4. X-Ray diffraction patterns for $Bi_{1.6}Pb_{0.4}Sr_2(Ca_{1-x}Sn_x)_2Cu_3O_\delta$ (LL-2201 phase, L-2212 phase, H-2223 phase and *-unknown phase)*

al. 1998) or formed as impurities. The existence of unidentified peak and low phase peaks contributed to the lowered value of $T_{C(R=0)}$ and T_{Conset} as compared to the pure sample. These results almost synchronised with the T_C values obtained from resistance measurement technique.

Fig. 5 showed the pattern of x-ray diffractograph for the samples doped with Sn in Cu site. In this doping mode, some of the high phase peaks were not observed and was found to have almost a similar pattern with the samples doped with Sn in Ca site. The degree of phase shifting in the peaks of H(0012), H(0014), L(115) and L(200) were bigger than that of the samples in the Ca doping mode. About five unidentified peaks appeared in this doping mode. The decreasing of Cu concentration as well as the increasing of Sn concentration introduced new peaks and hence lowered the superconducting properties.

The x-ray diffraction pattern for the samples doped with Sn simultaneously in Ca and Cu sites was shown in *Fig. 6.* It is clearly observed that some of the peaks disappeared and shifted in Sn doped samples. A new peak correspond to 2201 phase appeared in x=0.20 sample and was identified as LL(113). The unknown peaks still existed in this doping system.

The information of XRD pattern obtained for the three doping systems showed that the influence of Sn might cause the loss of some high phase peaks and helps in the formation of low phase peaks. The increasing of the volume fraction of low phase peaks contributed to the lowering of T_C value. It also decreased the strength of the coupling between the grains and increased the weak links behaviour. The shifting of the peaks probably caused the shortening

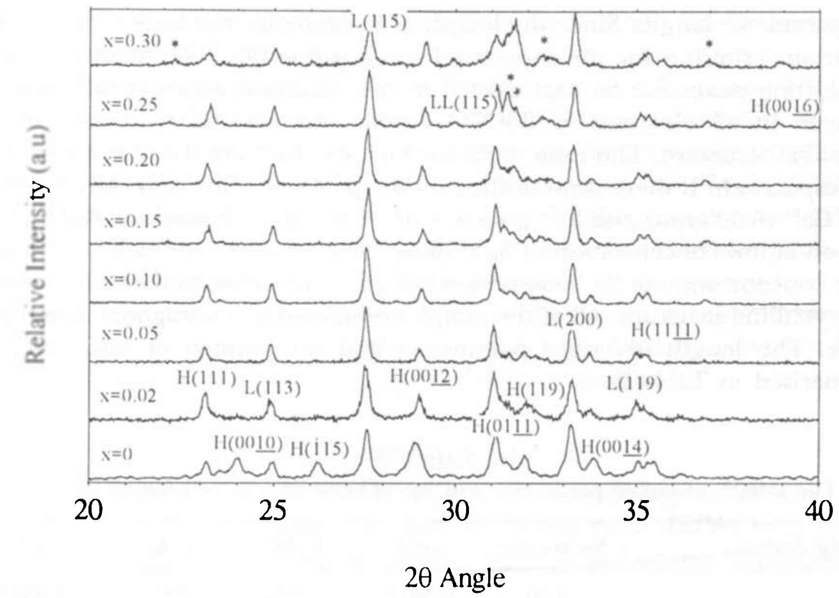


*Fig. 5. X-Ray diffraction patterns for $Bi_{1.6}Pb_{0.4}Sr_2Ca_2(Cu_{1-x}Sn_x)_3O_\delta$ (LL-2201 phase, L-2212 phase, H-2223 phase and *-unknown phase)*

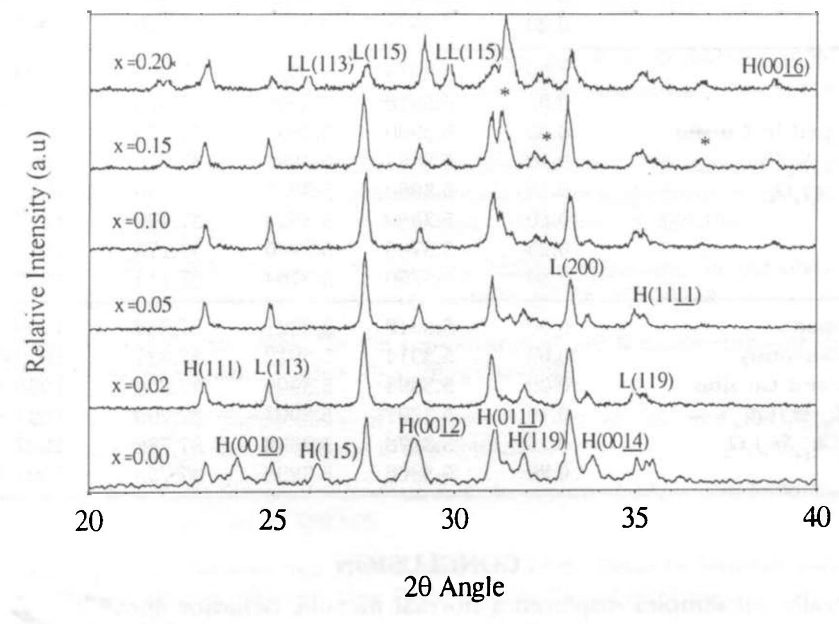


*Fig. 6. X-ray diffraction patterns for $Bi_{1.6}Pb_{0.4}Sr_2(Ca_{1-x}Sn_x)_2(Cu_{1-x}Sn_x)_3O_\delta$ (LL-2201 phase, L-2212 phase, H-2223 phase and *-unknown phase)*

of c-parameter length. Since the length of c-parameter was known to be a very important criterion for the superconducting properties. The main reason for the shifting peaks was Sn that existed in two oxidation states as Sn^{2+} and Sn^{4+} replaced in all elements in BSCCO system and was incorporated into the crystalline structure. The ionic radii for Sn^{2+} and Sn^{4+} are 0.093 nm and 0.071 nm respectively. If there is no shifting on the peaks, the Sn^{2+} may only substitute onto Ca^{2+} (0.099 nm) and Sn^{4+} goes to Cu^{2+} (0.072 nm). It was believed that this happen at low concentration of Sn doping such as x=0.02 and x=0.05. However if the concentration of Sn increased, it will go to the other elements and modify the crystalline structure. All of the samples remained in a tetragonal form which a=b≠c. The length of lattice parameters and the volume of unit cell were summarised in Table 2.

TABLE 2
The length of lattice parameter and the volume of unit cell for all samples

Doping System	Sn concen.	a(Å)	b(Å)	c(Å)	V(Å)^3
	0.00	5.3918	5.3921	37.837	1100.0
Sn doped in Ca site	0.02	5.3896	5.3897	37.836	1099.1
$Bi_{1.6}Pb_{0.4}Sr_2$	0.05	5.3890	5.3887	37.830	1098.6
$(Ca_{1-x}Sn_x)_2Cu_3O_\delta$	0.10	5.3888	5.3883	37.826	1098.3
	0.15	5.3881	5.3879	37.824	1098.1
	0.20	5.3878	5.3870	37.820	1097.7
	0.00	5.3918	5.3921	37.837	1100.0
	0.02	5.3910	5.3904	37.818	1099.0
Sn doped in Cu site	0.05	5.3900	5.3895	37.775	1097.4
$Bi_{1.6}Pb_{0.4}Sr_2Ca_2$	0.10	5.3884	5.3881	37.710	1094.8
$(Cu_{1-x}Sn_x)_3O_\delta$	0.15	5.3860	5.3865	37.506	1088.1
	0.20	5.3850	5.3853	37.490	1087.2
	0.25	5.3786	5.3780	37.175	1075.3
	0.30	5.3750	5.3764	37.114	1072.5
Sn doped	0.00	5.3918	5.3921	37.837	1100.0
simultaneously	0.02	5.3914	5.3912	37.827	1099.5
in Ca and Cu sites	0.05	5.3894	5.3909	37.804	1098.4
$Bi_{1.6}Pb_{0.4}Sr_2(Ca_{1-x}$	0.10	5.3891	5.3903	37.790	1097.8
$Sn_x)_2(Cu_{1-x}Sn_x)_3O_\delta$	0.15	5.3878	5.3898	37.786	1097.3
	0.20	5.3866	5.3808	37.753	1094.3

CONCLUSION

Generally, all samples displayed a normal metallic behavior above $T_{C onset}$. The values of $T_{C(R=0)}$ decreased towards Sn concentration. However the $T_{C(R=0)}$ value for x=0.02 sample doped simultaneously in Ca and Cu sites is 104 K which increased by 4 K as compared to that of the pure sample. In other systems of Sn doping (doped in Ca site Cu site) the $T_{c(R=0)}$ decreased at this concentration.

Samples doped with Sn in Cu site at x=0.20 and x=0.25 showed the lowering of resistive tail volume below the T_{Conset} as compared to the same value of Sn concentration in other doping systems. The decreasing of $T_{C(R=0)}$ and T_{Conset} with the increasing of Sn concentrations indicates that the substitution of Sn in BSCCO superconducting system did not favour for the formation of 2223 high phase but seems to be favourable to the formation of low phases (2212 and 2201). Sn as a non-magnetic element also reduced the coupling strength and hence increased the weaklinks between the grains. Some of the peaks have disappeared and shifted in the Sn doped samples. A new unidentified peaks and low phase peaks correspond to 2201 low phase exists at the level of Sn concentration above x=0.15. The existence of these particular peaks contributed to lower the superconducting properties in these systems.

ACKNOWLEDGEMENT

The financial support of the Ministry of Science, Technology and Environment Malaysia, under the Intensified Research in Priority Area (IRPA) vote: 4-07-05-026 (High Temperature Superconducting Materials and Device Applications) is gratefully acknowledged.

REFERENCES

NKUM. R. K. 1998. Superconductivity in (Bi, Pb)$_2$Sr$_2$Ca$_2$Cu$_{3-x}$V$_x$O$_y$. *J. Mater. Sci.* **33**: 207-210.

HALIM, S. A., S. B. MOHAMED, H. AZHAN, H. A. A. SIDEK and S. A. KHAWALDEH. 1999. Effect of barium doping in Bi-Pb-Sr-Ca-Cu-O ceramics superconductors. *Physica C* **312**: 78-84.

PISASS, M., D. NIARCHOS, C. CHRISTIDES and M. ANAGNOSTOU. 1990. The optimum percentage of Pb and the appropriate thermal procedure for the preparation of the 110 K Bi$_{2-x}$Pb$_x$Sr$_2$Ca$_2$Cu$_3$O$_y$ superconductor. *Supercond. Sci. Technol.* **3**: 128-133.

JHA, A. K. and R. G. MENDIRATTA. 1996. Effect of addition of barium, tin and silver on Bi-Sr-Ca-Cu-O ceramic superconductors. *J. Mater. Sci.* **31**: 1735-1740.

SESHU BAI, V. and S. RAVI. 1991. On the composition of 110 K superconductor in a (Bi, Pb)-Sr-Ca-Cu-O system. *J. Appl. Phys.* **70**: 4376-4382.

SANDEEP SINGH. 1998. Suppression of superconductivity in Sm and Co substituted Bi$_2$Sr$_2$Ca$_1$Cu$_2$O$_{8+\delta}$(system. *Physica C* **294**: 249-256.

NKUM, R. K. and W. R. DATARS. 1995. Weak links in ceramic In-doped Bi-Pb-Sr-Ca-Cu-O. *Supercond. Sci. Technol.* **8**: 822-826.

MELGHIT, K., S. RADHAKRISHNA and V. G. KUMAR DAS. 1998. *Trends in Materials Science*, ed. S. Radhakrishna, p. 181-193. New Delhi: Narosa Publishing House.

Preparation of Manuscript

General
The manuscript, including footnotes, tables, and captions for illustrations, should be typewritten double spaced on paper 210 x 297 mm in size, with margins of 40 mm on all sides. Three clear copies are required. Typing should be on one side of the paper only. Each page of the manuscript should be numbered, beginning with the title page.

Title page
The title of the paper, name of author and full address of the institution where the work was carried out should appear on this page. A short title not exceeding 60 characters should be provided for the running headline.

Abstract
Abstracts in Bahasa Melayu and English of not more than 200 words each are required for full articles and communications. No abbreviation should appear in the abstract. Manuscripts from outside of Malaysian may be submitted with an English abstract only.

Keywords
Up to a maximum of ten keywords are required and they should be placed directly below the abstract.

Footnotes
Footnotes to material in the text should not be used unless they are unavoidable. Where used in the text, footnotes should be designated by superscript Arabic numerals in serial order throughout the manuscript. Each footnote should be placed at the bottom of the manuscript page where reference to it is made.

Equations
These must be clearly typed, triple-spaced and should be identified by numbers in square brackets placed flush with the right margin. In numbering, no distinction is made between mathematical and chemical equations. routine structural formulae can be typeset and need not be submitted as figures for direct reproduction but they must be clearly depicted.

Tables
Tables should be numbered with Arabic numerals, have a brief title, and be referred to in the text. Columns headings and descriptive matter in tables should be brief. Vertical rules should not be used. Footnotes in tables should be designated by symbols or superscripts small italic letters. Descriptive materials not designated by a footnote may be placed under a table as a *note.*

Illustrations & Photographs
Illustration including diagrams and graphs are to be referred to in the text as 'figures' and photographs as 'plates' and numbered consecutively in Arabic numerals. All photographs (gloosy black and white prints) should be supplied with appropriate scales.

Illustrations should be of print quality; outputs from dotmatrix printers are not acceptable. Illustrations should be on separate sheets, about twice the size of the finished size in print. All letters, numbers and legends must be included on the illustration with the author's name, short title of the paper, and figure number written on the vereso. A list of captions should be provided on a separate sheet.

Unit of Measure
Metric units must be used for all measurements.

Citations and References
Items in the reference list should be referred to in the text by inserting, within parentheses, the year of publication after the author's name. If there are more than two authors, the first author should be cited followed by '*et al*'. The names of all authors, however, will appear in the reference list.

In the case of citing an author who has published more than one paper in the same year, the papers should be distinguished by addition of a small letter. e.g. Choa (1979a); Choa (1979b); Choa (1979c).

In the reference list, the names should be arranged alphabetically according to the name of the first author. Serials are to be abbreviated as in the *World List of Scientific Periodicals.*

The abbreviation for Pertanika Journal of Science and Technology is *Pertanika J. Sci. Technol.*

The following reference style is to be observed:

Monograph
Alefed, G. and J. Herzberger. 1983. *Introduction to Interval Computations.* New York: Academic Press.

Chapter in Edited Book
Muzzarell, R.A.A. 1980. Chitin. In *Polymers in Nature,* ed. E.A. MacGregor and C.T. Greenwood, p. 417-449. New York: John Wiley.

Serials
Kamaruzaman Ampon. 1991. The effect of attachment of hydrophobic imidoesters on the catalytic activity of trypsin. *Pertanika* **14(2)**: 18-185.

Proceedings
Mokhtaruddin, A.M. and L.M. Maene. 1981. Soil erosion under different crops and management practices. In *Agricultural Engineering in National Development,* ed. S.L. Choa, Mohd Zohdie Bardaie, N.C. Saxena and Van Vi Tran, p. 245-249. Serdang, Malaysia: Universiti Pertanian Malaysia Press.

Unpublished Materials (e.g. theses, reports & documents)

Sakri, I. 1990. Proper construction set-up of Malaysian Fish Aggregating Devices (Unjam). Ph.D. Thesis, Universiti Pertanian Malaysia, Serdang, Selangor.

Pertanika Journal of Science & Technology

Volume 10 No. 1, 2002

Contents

ISSN 0128-7680